A computational and psychophysical study of motion induced distortions of perceived location

by

Szonya Durant

Submitted to the Centre for Mathematics and Physics in the Life Sciences and Experimental Biology and Psychology Department
University College London

in partial fulfilment of the requirements for the Degree of

Doctorate of Philosophy

January 2004

Abstract

In this thesis I begin by extending previous psychophysical research on the effects of visual motion on spatial localisation. In particular, I measured the perceived spatial shift of briefly presented static objects adjacent to a moving stimulus. It was found that the timing of the presentation of static objects with respect to nearby motion was crucial. I also found a decrease of this motion induced spatial displacement with the increasing distance of static objects from motion, suggesting a local effect of motion. The induced perceptual shift could also be reduced by introducing transient stimuli (flickering dots) in the background of the display.

The next stage was to construct a computational model to provide a mechanism that could facilitate such shifts in position. To motivate our combined model of motion computation and spatial representation we considered what functions could be attributed to V1 cells on the basis of their contrast sensitivity functions. I found that functions based on sums of differential of Gaussian operators could provide good fits to previously found V1 data.

The properties of V1 cells as derivatives of Gaussian kernel filters on an image were used to build a spatial representation, where position is represented in the weighting of these filter outputs, rather than in a one-to-one isomorphic representation of the scene. This image representation can also be used along with temporal derivatives to calculate motion using the Multi-Channel Gradient Model scheme (Johnston et al, 1992). I demonstrate how this framework can incorporate motion signals to produce “in place” shifts of visual location. Finally a combined model of motion and spatial location is outlined and evaluated in relation to the psychophysical data.

Acknowledgements

This work was supported by a Medical Research Council Bioinformatics/Neuroinformatics PhD Studentship awarded by the Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX) and completed in collaboration with the Psychology Department at University College London.

I would like to thank my main supervisor Alan Johnston at the Psychology Department, University College London. In particular I would like to thank Alan for his incredible support and enthusiasm throughout and super-speedy proof reading. I would also like to thank my second supervisor Jaroslav Stark from the Mathematics Department at Imperial College London.

I would like to thank Glyn Cowe, Jason Dale and Derek Arnold for help with programming, explaining and advising, and Hazel Savage for some of the data collection. Finally I would like to thank my friends Essie and Jo for last minute proof reading, my parents for their constant emotional support and most importantly my boyfriend Drew for helping me throughout.

Contents

Chapter 1- Background

Computational models of visual motion processing have traditionally involved hierarchical structures. Such models are constructed from basic spatio-temporal filters, combined to form directionally specific motion detectors, which in turn are combined to produce more complicated velocity computation modules. Many successful algorithms that follow this hierarchical structure have been developed that can extract motion from a scene and calculate its direction and magnitude (Adelson & Bergen, 1985; Johnston et al., 1999; Zanker, 1994).

Popular biologically plausible motion detection models are based on physiological evidence of properties of visual neurons along the visual pathway. Past models incorporate the properties of cells in the retina and lateral geniculate nucleus (LGN), those of the primary visual cortex (V1), and finally encompass the motion specialisation of cortical area V5/MT. The discovery of V5/MT as the specialised area for motion (Tootell et al., 1995; Zeki, 1978) detection has led researchers to reconstruct the motion pathway in the brain as a separate motion detection machine. However, accurate motion detection does not by itself provide us with all the visual information we need to survive. Therefore the human brain has evolved with the motion detection system as an integrated part of the complete system. Motion information is combined with spatial information and further aspects of the scene over time to provide the whole picture.

It is perhaps not surprising to find that motion and spatial position are in fact inextricably linked. Increasingly, empirical evidence suggests the motion in the visual scene can affect the way in which we see the position of objects. This is an interesting and rich area for research not only because of the new slant it gives on traditional motion processing models. The question of motion-spatial position interaction also calls for a re-examination of the thorny philosophical

issues of localisation in space and the possible neural reconstruction of a scene. Finally, we touch on how two aspects of the visual scene are combined over the time of a neural response. This involves challenging both the feed-forward, hierarchical nature of motion processing models and the naïve assumption of point-wise spatial representation that often still underpins the question of visual localisation.

At some point a simple one-to-one correspondence between stimulus and percept breaks down in the visual pathway as is shown by the existence of any visual illusion, where the final percept differs from the image formed on the retina. In particular, I will describe several perceptual effects where there is a shift in local position between percept and stimulus. I propose that these illusions arise at the V1 level; it is here that the direct correspondence between retinal image and brain activity breaks down. This is contrary to the usual assumptions. In the past, as V1 has been found to contain a retinotopic map V1 (Daniel & Whitteredge, 1961; Bosking et al., 2002; Tootell et al., 1998), it has been assumed that position in V1 is represented by the firing of a specific cell or population of cells, i.e. the firing rate of cell A represents luminance at position A. These kind of schemes originate with the idea of 'local sign', with the associated question of how such a 'place label' gets attached to a neuron (Koenderink, 1984; Lotze, 1884). Within the scheme proposed in this thesis, position information is represented by the combined activity of V1 cells. The same set of V1 cells fire under the influence of visual motion as without, but their combined output is altered so that the overall output is equivalent to a positional shift (see Fig. 1.1).

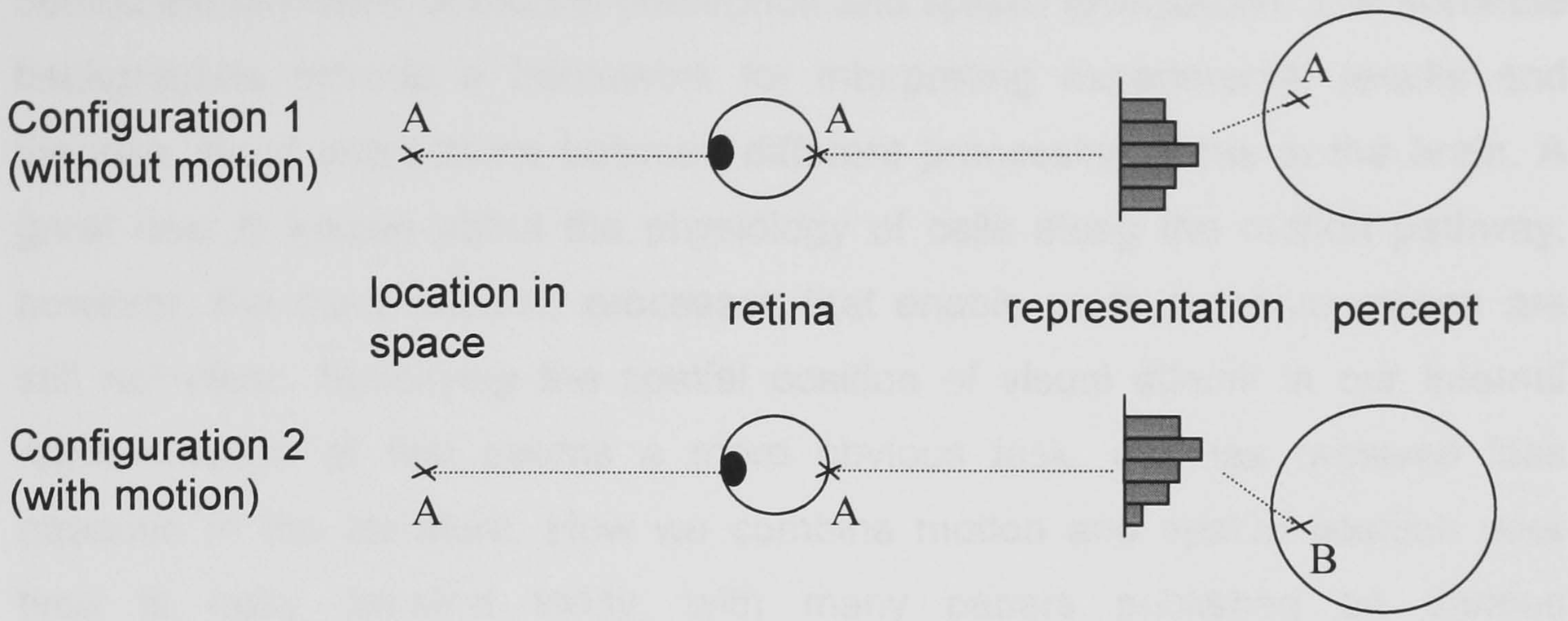


Fig. 1.1 There is not always a direct correspondence between retinal image and the percept formed. It is suggested that the way the representation is formed can be altered by motion.

In order to model translation effects we need to be able to alter the representation of the position of the stimulus to produce the effect. It seems reasonable to suggest that visual motion affects the way in which a spatial representation is formed.

In this thesis I will extend the empirical investigation into the properties of the motion induced shift to demonstrate that it is a local effect that evolves during the neural response to a static stimulus. I then propose a model of V1 neural response that can accurately describe single cell behaviour as well as provide a labile spatial representation that can produce in-place shifts of position caused by the input of visual motion. I aim to model the perceived positional shift as the result of a feedback loop from the motion calculation area of V5/MT to the accurate spatial representation in V1.

In this chapter, first of all, we will review the physiological, psychophysical and computational work that informs the current view of motion and spatial position as separate aspects of the visual scene. I will then present an overview of the evidence that motion processing interacts with the way in which humans spatially locate objects in the scene. Finally, I bring together the theories that attempt to answer how and why we might combine these aspects.

The first section describes some of the physiological and computational ideas behind explanations of motion perception and spatial localisation. The separate backgrounds provide a framework for interpreting experimental results and theories about interactions between different processing areas in the brain. A great deal is known about the physiology of cells along the motion pathway; however, the computational processes that enable us to perceive motion are still not clear. Specifying the spatial position of visual stimuli in our internal representation at first seems a more obvious task, but has received less attention in the literature. How we combine motion and spatial position over time is hotly debated today, with many papers published on various phenomena, but no consensus formed.

1.1 - V1 receptive fields

Each cell along the visual pathway has its own spatial receptive field – an area on the retina that when illuminated will cause the cell to fire or inhibit the cell from firing. It is more properly referred to as the classical receptive field, to differentiate direct from indirect excitation of the cell. It is useful to describe receptive fields in terms of excitatory and inhibitory regions. These regions can be hand plotted by recording responses of a cell over its receptive field to a spot of light. In the retina and lateral geniculate nucleus (LGN) these take on a simple circular symmetric centre-surround shape, with a central excitatory region flanked by a surrounding inhibitory region (or vice versa) (Barlow, 1953; Kuffler, 1953). These early cells produce the largest responses when the excitatory regions are illuminated and inhibitory regions are in the dark, avoiding the cancelling out that equal excitation and inhibition would cause under uniform lighting. Uniform light produces a low response and differences in brightness within the receptive field cause maximal firing. Not only do these cells seem to be filtering out information such as uniform light, but also if light changes occur often enough over the space within the receptive field, then the inhibitory response will cancel out the excitatory again. Thus these cells can

also be characterised as band-pass spatial frequency filters, as they only respond to change in light intensity at a certain range of spatial frequencies (De Valois & De Valois, 1990).

One of the main cortical areas of interest to us, area V1, contains cells further on in the vision hierarchy. Cells from the lateral geniculate nucleus feed into V1 cells in such a way as to give different, more complicated properties to these cells. The magnocellular layers in LGN respond best to changes in luminance and are characterized by their transient responses, and the parvocellular layers in the LGN have a more constant response level to permanently presented stimuli (Derrington & Lennie, 1984). V1 cells tend to be classified into two groups: simple and complex cells. Simple cells are defined by the fact that they exhibit linear summation, whereas complex cells have non-linear properties (Hubel & Wiesel, 1962, 1968). Most V1 cells respond best to long bars of light. In particular, a V1 cell will usually respond best at a certain orientation of the bar. Other V1 cells are the reverse and respond well to dark strips of a certain orientation. Other V1 cells respond maximally to a straight edge between a light and a dark area. This description assumes a fixed contrast; varying contrast will also affect cell response, making single cell responses ambiguous. Receptive fields for simple cells can be mapped out as before, by hand-plotting excitatory and inhibitory regions, which in this case are clearly defined. As expected, the shape of the fields corresponds to the different kinds of effective stimuli (Hubel & Wiesel, 1962). For light bar, dark bar and edge detectors we find a long narrow excitatory region flanked by two larger separated inhibitory regions, the same in reverse, and an excitatory region with a clear straight boundary next to an inhibitory region, respectively.

The non-linear properties of complex cells are caused by the absence of clear 'on' and 'off' regions. Complex cells also have a preferred orientation, but they are less sensitive to where the bar of the correct orientation is presented in their receptive fields (Hubel & Wiesel, 1962). In this way, complex cells are considered to be phase insensitive e.g. they will fire at constant rate to a drifting

sine grating presented in their receptive files, whereas simple cell response will modulate in phase with the sine grating. Often, however, cells in V1 don't fit exactly into either of these categories. They can be classified more quantitatively by measuring cell responses to static sine gratings over various phases. Simple cells are defined as cells whose ratio of the mean of the second harmonic responses over the peak amplitude of the fundamental response is less than 1, that is cells with a dominant linear component. Many V1 cells, both complex and simple, also have the additional property of responding better to a certain direction of movement than other directions (De Valois et al., 2000; Emerson & Gerstein, 1977; Hubel, 1995; Hubel & Wiesel, 1968).

Like spatial receptive fields, the temporal receptive fields of cells (i.e. the pattern of excitatory and inhibitory response of a cell over time) can also be mapped out. It has been found that in the non-directionally selective V1 cell population there are two distinct sub-populations (De Valois et al., 2000); those with slow, largely monophasic temporal receptive fields and those with a fast biphasic response, which in turn may correspond to cells in the P and M pathways respectively. De Valois et al. (2000) have found through principal components analysis that the temporal receptive fields of directional cells could be constructed by a linear combination of these two components.

It is useful to find mathematical functions that describe the shape of 2D spatial receptive fields as these can be used to predict the sort of transformations these patterns of excitation and inhibition could accomplish. Predictions are generated by convolving mathematical functions with an image. We can reduce the problem to 1D for V1 cells, by finding the optimal orientation of the cell and plotting its response to light orthogonally to this axis. For V1 simple cells this will produce a class of recognisable curves.

Three main types of descriptions have been used in the past and these are Gabor (wavelet) functions, the second derivative of a Gaussian or the difference of two Gaussians. In Chapter 3 we will describe each of these functions in detail. These functions, in turn, have then been used as parts of

algorithms for extracting edges, finding orientation, spatial frequency and the amount of blur in a scene (Georgeson, 1991; Marr & Hildreth, 1980; Sherwood & McOwan, 2003). The approach taken in this thesis is to build a model that is consistent with existing physiological data. Accordingly a model of spatial representation and motion detection should have filter inputs that behave in the same way as V1 cells and therefore the functions used must provide good fits for V1 cell receptive field patterns.

1.2 - Motion detection

Motion is the physical change of position over time. Motion detection by the visual system, however, is not simply a by-product of the perceptual position and timing of an event. Motion seems to be a separate attribute processed in parallel. The motion aftereffect (MAE) provides a simple demonstration of the separate nature of motion. On being presented with a static pattern after adapting to a moving pattern, we experience a feeling of motion in the opposite direction to the adapting stimulus. However, there is no change in the physical position of the static stimulus and we do not perceive the same change in position we would for a physically moving stimulus. It seems that the motion percept exists separately from perceived change in position. Exner (1888) demonstrated the primary nature of the motion percept by causing two sparks successively so spatially close to one another that an observer could not distinguish them in space. Despite their inability to spatially resolve the sparks, the observers reported a sensation of motion. He also demonstrated that motion is experienced when the observer is unable to temporally resolve the sparks.

The first cells along the primate visual pathway that are selective for motion are in V1, the primary visual cortex, where some simple cells and most complex cells respond optimally to a certain direction of motion and some respond optimally to a certain speed of motion (De Valois & De Valois, 1991; Hubel,

1995). The V5/MT region mentioned above is an area outside V1 that is specialised for motion. V5/MT cells have larger receptive fields than V1 cells and are unresponsive to static pattern (Zeki, 1978). The specificity of this area was further emphasised through PET (positive emission tomography) (J. D. Watson et al., 1993) and fMRI (functional magnetic resonance imaging) (Tootell et al., 1995) studies. Neuropsychology in turn has provided us with the example of patient LM, who was found to have a bilateral lesion located in the lateral temporal-occipital cortex. This patient has a specific visual deficit. She reports seeing consecutive "snapshots" rather than smooth motion (Zihl et al., 1983). High resolution MRI of her brain showed that the zone occupied by area V5 had indeed been destroyed bilaterally (Shipp et al., 1994). Further testing of the patient revealed that residual motion perception was associated with area V3A (Shipp et al., 1994). This kind of concentrated selectivity also demonstrates that motion is not simply a by-product of space-time calculations, but an evolutionarily important component of the visual scene in its own right. The motion pathway is also thought to be one of the fastest as the relative latency involved in detecting the direction of movement has been found to be shorter than for other visual tasks (Allik & Kreegipuu, 1998).

The discovery of a motion specific area outside V1 has meant that motion perception has been studied as a completely separate process. This modular view of the brain has been strengthened by the discovery of other specific areas, such as V4 for colour (Zeki, 1978). However, it has also been known that for the projections of V1 onto V5/MT there are an equivalent number of recurrent connections (Ffytche et al., 1995; Shipp & Zeki, 1989). Increasingly it is becoming clear that early processing at the V1 level may interact with other processes. The possibility that some kind of V1-V5/MT feedback loop exists has been raised, altering the previous hierarchical view (Bullier, 2001a; Pascual-Leone & Walsh, 2001; Shipp & Zeki, 1989).

Cells in V5/MT have larger receptive fields than V1 cells at corresponding eccentricities (Van Essen et al., 1981). This may reflect a summation over

space to help to compute motion more globally rather than locally. This would help solve the aperture problem (Fig. 1.2) posed by motion selectivity of small receptive fields. The aperture problem refers to the fact that the movement of a border seen through a small aperture is ambiguous in direction. Motion is seen perpendicular to the edge.

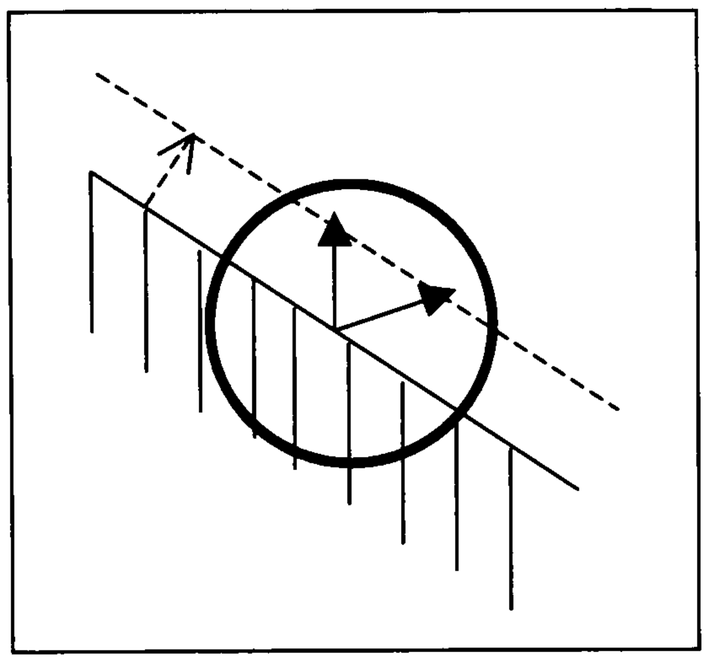


Fig 1.2 The aperture problem. Using information only from within the circular receptive field, either of the two motion directions shown is possible.

Hence it might be necessary to combine small localised spatial receptive fields as found in V1, into larger ones, to resolve such ambiguities over the whole of the scene.

In understanding motion detection we need to combine different levels of the visual system. It has been suggested that retinal ganglion cells in humans behave as spatio-temporal filters and these can be combined to produce the direction specific properties of V1 cells. This computational approach has given rise to three main types of motion detection algorithms, which can overlap according to the specifications of their components. These three types of models are the correlation (Reichardt) detector based model (Reichardt, 1961; Zanker, 1994), motion energy models (Adelson & Bergen, 1985; Chubb & Sperling, 1989) and gradient models (Horn & Schunck, 1981; Johnston & Clifford, 1995). We will discuss an image as a 2D luminance signal over time, which by Fourier analysis can be decomposed into cosine and sine waves of

varying spatial and temporal frequency. When building mathematical models of the visual system there are two main considerations: to base models on what is known from physiology; and to make models as effective and efficient as possible. It is useful to think of motion as an orientation in space-time. If we consider only one spatial direction and plot this against time, then a trajectory of constant velocity would appear as a diagonal line. The orientation of this line tells us the velocity of the point moving along the trajectory.

The first step in building a motion detection mechanism is to combine non-directionally selective inputs that are separated from each other in time and space, using their product as a motion indicator. In this way non-oriented filters are combined to form spatio-temporally oriented filters for the detection of motion.

The simplest example of combined non-directional inputs forming a direction selective output is a delayed product mechanism. See Fig. 1.3.

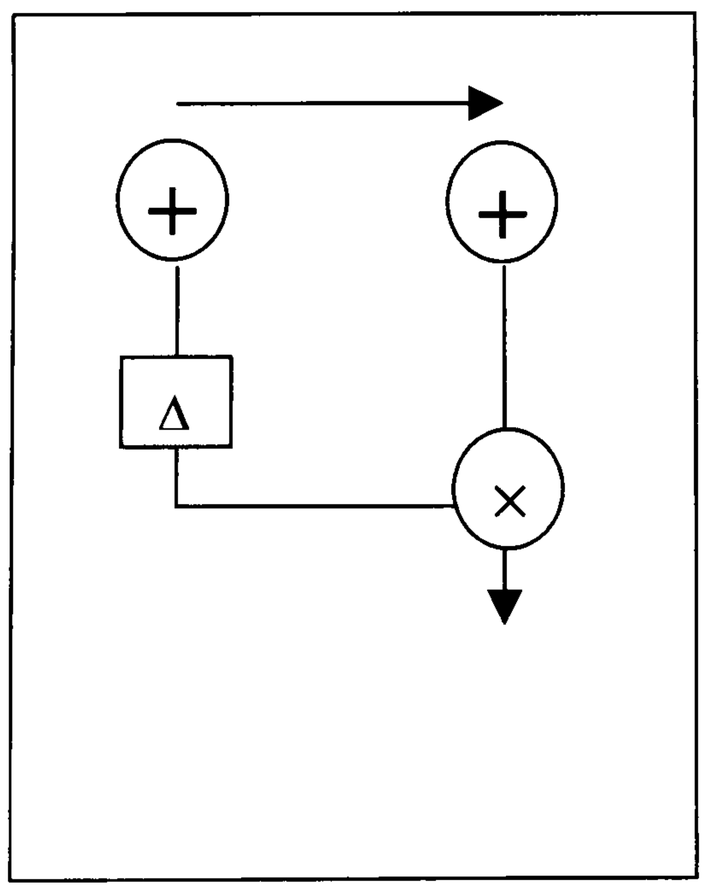


Fig. 1.3 Simple Reichardt detector. "+" represents excitatory input. Δ is a small temporal delay.

We can see that by delaying the response of one of the filters, we detect the product of the impulses together in the preferred direction. By combining an excitatory and inhibitory filter in a similar way, we can achieve a null response in one direction and a maximal response in the opposite direction (Barlow & Levick, 1965). These simple mechanisms were not initially developed for human vision; for example, delayed inhibition is a model of the response properties of rabbit retinal ganglion cells (Barlow & Levick, 1965). Human vision motion models also use non-directional inputs to build motion filters. We will briefly examine the three main types of models here.

1.2.1 The Reichardt correlator model

This model uses leftward and rightward motion sensitive filters that have been constructed in a similar way to those described above. It uses the multiplication of spatially and temporally separated filters to form the motion selective units and the output is then the overall difference between rightward and leftward motion detectors. This model is generally contrast sensitive and in many cases single motion detector components cannot calculate velocity. Hence a further scheme often needs to be included for combining several differently tuned velocity detectors to calculate an overall velocity (Borst & Egelhaaf, 1989). However, it has been shown that under certain conditions single unidirectional motion detector can produce responses that vary with image speed (Zanker et al., 1999). Also, these models are relatively simple to simulate and can be powerful in predicting various psychophysical phenomena (Borst & Egelhaaf, 1989; Zanker, 1994). For instance, the Reichardt type model provides an intuitive explanation of the MAE; adapting to rightward motion causes the rightward detectors to signal less over time, so, given a static stimulus after adaptation if we subtract rightward motion from leftward motion we get an overall signal of motion to the left. This model is called the correlation model as it finds correlation between signals over space and time to find the direction of motion for which there is the most agreement.

1.2.2 Motion energy

This type of model creates direction-selective filters by adding and subtracting the responses of non-directional filters. Filters oriented in space-time are paired in quadrature i.e. they are 90° out of phase with each other. The outputs are then squared and added for each quadrature pair, to give rightward and leftward 'motion energy' over all pairs. The opponent stage takes the difference of the two energy values to eliminate responses to flicker. To encode velocity without being affected by contrast, motion energy needs to be divided by a measure of "static energy" (Adelson & Bergen, 1985; Chubb & Sperling, 1989). In this way opponent energy can be used to find a velocity measure. The squaring and summing of inputs produces phase invariance – this represents complex cell properties in V1. It has been argued that the generic model is still sensitive to contrast and may also be affected by static pattern (Johnston & Clifford, 1995).

1.2.3 Gradient models

Marr, in his book *Vision*, suggests that the parvocellular system works to find gradients of light intensity over space for use in detection of edges for example. He also suggests that, based on the neurophysiological recordings of magno cells, their transient responses can be used to measure temporal derivatives (Marr, 1982). Gradient based methods calculate motion by using spatio-temporal gradients of light intensity (Horn & Schunck, 1981; Johnston et al., 1999; Johnston et al., 1992). There is some psychophysical evidence that we do perceive motion by combining spatial and temporal gradient detectors. It has been shown that adapting to a spatially uniform patch, ramped in brightness over time, then viewing a spatial luminance ramp at the same location, causes a percept of directed motion. The ramping of luminance over time will not cause differential adaptation in spatio-temporally oppositely oriented detectors, so there should be no effect of adaptation if these form the basis of motion perception (Anstis, 1990).

If we only consider change in one spatial direction, speed is given by the ratio of derivatives of intensity with respect to time and space:

$$v = \frac{\partial I}{\partial t} \Big/ \frac{\partial I}{\partial x} \tag{1.1}$$

This basic notion is incorporated in all gradient models, including the Multi Channel Gradient model (McGM) (Johnston & Clifford, 1995; Johnston et al., 1999; Johnston et al., 1992). It is shown that these gradients can be found with the use of two spatially overlapping, contiguous, space-time separable filters – a transient temporal filter and a sustained spatial filter – which may in turn correspond to the transient and sustained properties of M and P cells. However, the problem with equation 1.1 occurs when there is no change in the image intensity over space, i.e. $\frac{\partial I}{\partial x}$ = 0. This is overcome by taking the Taylor approximations for average brightness and taking the spatial and temporal derivatives of all of the separate terms to form two vectors x and t. An approximation of v can be expressed as:

$$v = \frac{\mathbf{x} \cdot \mathbf{t}}{\mathbf{x} \cdot \mathbf{x}} \tag{1.2}$$

This expression is only ill-conditioned if all spatial intensity values are equal to zero, in which case the concept of image motion no longer makes sense. This model recovers an accurate measure of velocity. The extended motion energy model can strongly resemble the gradient model. However, this depends on which components we choose to construct the oriented filters in the energy model. I will present arguments that there is physiological evidence for derivative type filter shapes for V1 cell responses (see Chapter 3) and go on to describe the McGM scheme in more detail (see Chapter 4), along with possible adaptations for modelling motion-spatial location interaction (see Chapters 5 & 6).

One can aim to build on existing models for computing velocity in order to generate an integrated model that can also account for the processing of spatial and temporal information.

1.3 - Spatial localisation

A question perhaps less often examined than “How do we detect motion?” is that of how we maintain the precise spatial mapping of light offered up by the retina to higher levels.

One measure of our sensitivity to spatial variation is visual acuity. Psychophysical measurements of acuity show that as inter-receptor spacing increases, visual acuity correspondingly decreases as one moves towards the periphery (Westheimer, 1981). The spatial separation of photo-receptors defines the upper limit of acuity. In turn cortical receptive field size is related to cortical spatial sampling intervals and hence smaller receptive fields imply higher position acuity. The small retinotopic receptive fields in V1 make it a plausible brain area for the representation of positional information.

However, visual acuity tasks such as judging whether one or two lines are presented have been found to result in less accurate performance than the task of aligning two separate line segments (Morgan & Ward, 1985; Westheimer, 1981). This measure of visual precision is called vernier acuity and this task is done more accurately than the sampling density of the visual system would appear to allow. This is an example of hyperacuity and measures the observer’s ability to localise a stimulus, rather than the ability to resolve spatial qualities. It has been suggested that this is achieved by making use of the blurring of light from the stimulus and a simple differencing operation (De Valois & De Valois, 1990). This task has been exploited to try and find out whether we make use of an explicit spatial representation of target feature position (Morgan et al., 1990). However, it has also been argued that relative spatial position is implicitly encoded by spatial frequency channels (Georgeson, 1980).

If we do in fact explicitly encode position, then we are faced with the problem of local sign or *Localsheizen* as described by the German philosopher Lotze (1884). He questioned whether it was spatial position in the brain that implied position or whether it was the functional position in a network that was important. Hence "labelled lines" would signal position, but it is not clear how such a place label or local sign might become attached to an optic nerve fibre. Koenderink (1984) tackled this problem by suggesting that position could be encoded in the logical relationships between the firing patterns of cells over time. For instance, he proposed that if the activity in two neurons was correlated then you could suspect proximity between the two. In this view "local sign" cannot be built in, it can only be acquired. This raises the idea that position may not be simply coded by one cell, but rather be inherent in the overall pattern of cell activity.

We can also try to address the vernier alignment task along the lines of how we might apply spatial filters to derive the difference in position. Vernier acuity could in principle be accounted for by orientational specificity without making spatial position explicit. Changing the spatial separation between two dots would change the overall activity between orientationally tuned channels and also across size-tuned spatial frequency channels. Accordingly, models using orientation specific channels have been suggested (Sullivan et al., 1972) along with spatial frequency based localisation models (Wilson & Gelb, 1984).

We know we have a retinotopic mapping onto V1 (Daniel & Whitteredge, 1961; Bosking et al., 2002; Tootell et al., 1998), which could be considered useful when locating visual objects. As well as considering the position information given by the position of a cell within V1, spatial information may be gained from within the receptive fields of cells. If motion can be described as changing temporal and spatial frequencies, then spatial localisation within the receptive field of a cell can be considered as reflecting the phase of these luminance waves. Some retinal ganglion cells have been found to be phase sensitive, whereas others have not (De Valois & De Valois, 1990). Cells along the

parvocellular pathway respond well to permanent stimuli, so they give a constant signal when a stimulus is present at a certain location. We have also already mentioned that simple cells in V1 are sensitive to where a correctly oriented bar is presented in their receptive field, whereas complex cells will respond similarly to a correctly oriented bar at any position in their receptive field (Hubel, 1995).

Representing spatial position has often been thought of in terms of encoding spatial pattern. This can be done through Fourier analysis of luminance waves. Analysing spatial pattern in this way tells us about relative position, but not absolute position. V1 cells are often thought of as local spatial frequency analysers (Georgeson, 1980). Possible models utilise the fact that simple cells with odd-symmetric receptive fields and simple cells with even-symmetric receptive fields exist in the cortex. These cells are 90° out of phase with respect to each other (in quadrature). If these cells were aligned so that the centre of their receptive fields coincided then we could in theory encode the spatial phase by their relative activity rates (De Valois & De Valois, 1990).

Our most accurate spatial representation exists at the V1 level, yet at this level we require the same set of cells to give us information about different aspects of the scene. This multiplexing means that it is not clear how to separate spatial position from other properties of components of an image. It also seems that this precise absolute positional information is not necessarily retained in higher cortical areas. Small local receptive fields by themselves can lead to ambiguities as we have seen from the aperture problem (Fig 1.2). The question arises of how we maintain accurate positional information at the same time as aggregating responses to acquire global percepts.

1.4 - Interaction

We have discussed motion detection and spatial localisation as separate processes. However, there is evidence that, in addition to breaking the scene into these separate parts, interaction also occurs. Recent experiments have led to the idea that the very action of processing motion might interfere with our ability to discern spatial and possibly temporal attributes of the stimuli. This is especially interesting as motion processing involves higher cortical areas outside V1, whilst it seems likely that accurate spatial localisation needs to take place in V1. It also has been a popular concept that we have separate visual pathways for localisation (P pathway) and motion detection (M pathway). This is because, as we have described, P cells have more sustained responses, as opposed to the transient responses of M cells. The following experiments have tried to address what happens when we need to access combined information.

Interactions between motion and spatial processing could simply be caused by differences in the visual pathways such as relative processing times for example. There is also the possibility of low-level interactions and feedback loops. In this section we are going to examine the empirical evidence that shows different effects of interaction, the parameters that change the size of these effects and proposed models that incorporate the evidence.

1.4.1 The effect of brightness and contrast

It has been known for a long time that the relative perceived position of moving objects can be influenced by their relative brightness. In a classic experiment, Hess showed that a bright line translating across space in alignment with a darker bar appears to be spatially more advanced (C. V. Hess, 1904). In a related example, the Pulfrich effect (Pulfrich, 1923) is seen when a pendulum is viewed with both eyes, one of them covered with a neutral density filter. This creates the illusion that a pendulum swinging in a plane perpendicular to the line of view of the observer is rotating in depth. This is because lowering the

contrast in one eye makes the response in that eye lag behind, causing a disparity that leads to illusory depth. These experiments have been considered an illustration of neural latency causing the misjudgement of relative positions of moving objects. The brighter an object the quicker photoreceptors respond, hence the information from the bright object will travel more quickly through the visual system than the dimmer one. In this case, however the neural latency difference is caused by introducing latencies at the retina. Later latency difference explanations implicate in-built relative processing delays in the cortex. (See section 1.4.4) Related to these brightness effects is the "flash-lead" phenomenon (the opposite to "flash-lag", which we describe below, see section 1.4.4) (Patel et al., 2000). It was found that by presenting a brighter bar flashed in line with a dimmer moving bar, they could make the flash appear to be spatially ahead of the moving bar. It was argued that the increase in light intensity causes the flash information to reach a saliency threshold quicker and hence awareness of its position occurs before awareness of the position of the moving object.

1.4.2 Start and end position along a trajectory

Further examples of misjudging the position of moving objects can occur at the start and end of a moving object's trajectory. When subjects are asked to determine where a fast-moving stimulus enters a window, they typically do not localize the stimulus at the edge, but at some more spatially advanced position within that window (Musseler & Aschersleben, 1998). This is the Frohlich effect, which is usually explained in terms of attention. Representational momentum (Hubbard, 1995) is the corresponding effect at the end of a trajectory, where the final position of a moving object is judged to be further ahead than presented. Representational momentum experiments typically involve presenting a series of object positions taken from a motion sequence before the test target. The test target position is judged to be advanced in the direction of motion. However, this effect is found even with implied motion, where the ISIs are so long that no motion is perceived. Both these effects have also been

investigated as the flash-initiated and flash-terminated flash-lag paradigm (see section 1.4.4), with no forward shift found for the representational momentum case (Eagleman & Sejnowski, 2000).

1.4.3 Perceptual deformation induced by visual motion

It has also been found that a difference in position can be observed between physically aligned moving components, caused simply by the different positions the components take over the trajectory. This difference in position deforms the overall shape of the pattern formed by the moving objects. Matin et al. (1976) showed that a line segment rotating in line with two separate line segments either end caused a vernier misalignment. This could be due to the different speeds of the two lines or their different eccentricities on the retina. A similar effect was shown with three vertically aligned drifting dots, translating horizontally (Zanker et al., 2001). The middle dot appeared to be leading the other two dots. In this case the dots had all the same speed and were not sufficiently separated to make differential processing latencies a possibility. The size of this deformation increased with greater speeds. It was suggested this could be due to spatial blur emphasising the luminance of the central dot and leading to a similar case as above, in which brighter objects appeared more advanced.

1.4.4 The flash-lag effect

The flash-lag illusion (MacKay, 1958; Nijhawan, 1994) is a classic motion effect that can be explained by the mislocalisation of objects or mistiming of events caused by the presence of motion. More specifically it arises from the comparison of moving objects with static ones. It describes the fact that a briefly presented static object appears to spatially lag behind a moving object. For example a static bar flashed briefly in line with a moving bar appears to be behind the moving bar (see Fig. 1.4).

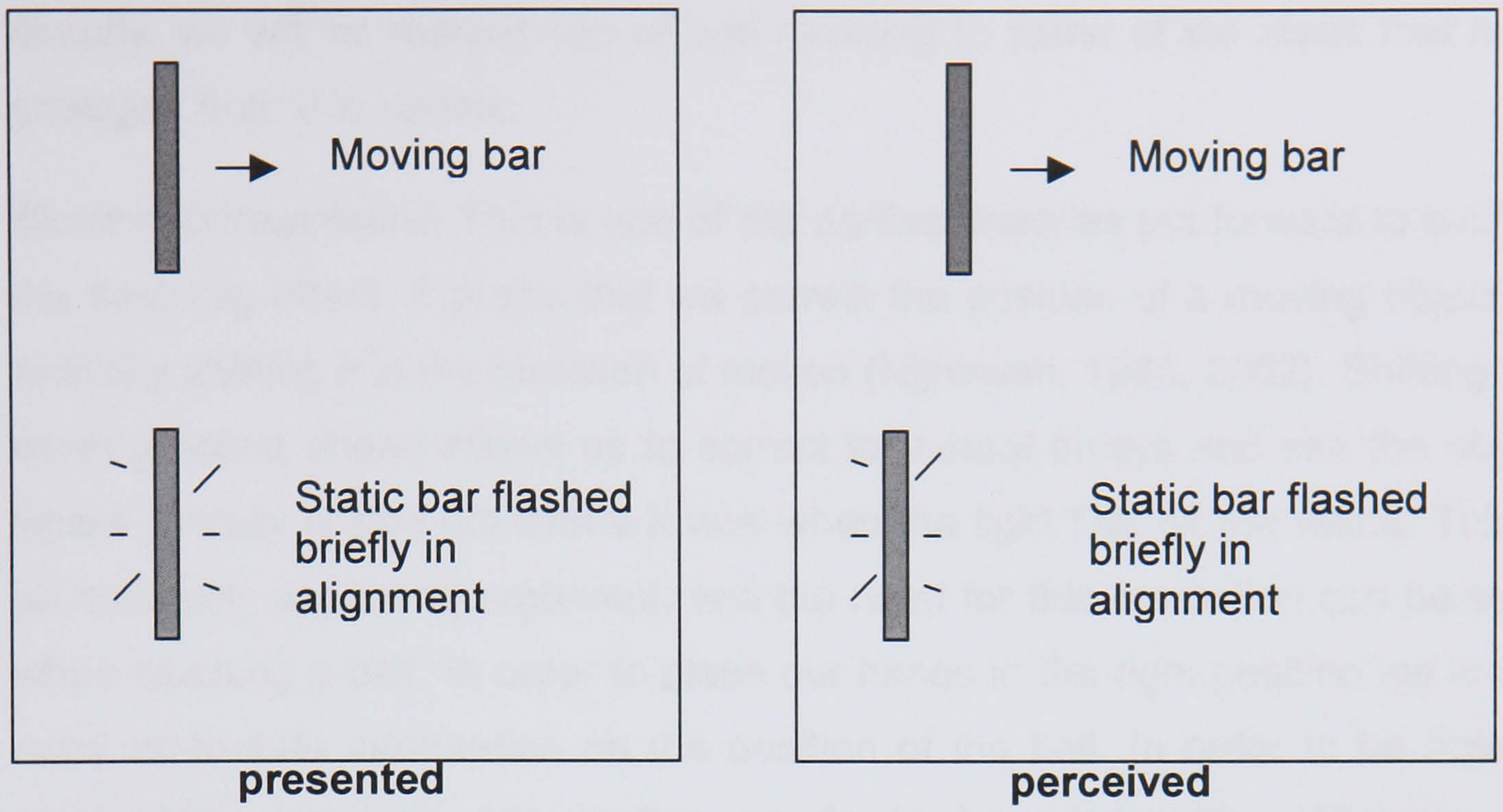


Fig. 1.4 The flash-lag effect. The relative position of a moving bar and a briefly presented static bar is misperceived, so that the static bar appears to spatially lag behind.

Nijhawan (1994) demonstrated the effect using a rotating bar with flashing ends. The ends of the bar appear to lag behind the middle. Whitney et al. (2000) established that the same effect still existed when motion was not predictable. When the moving bar suddenly changed direction, the flashed bar still appeared to lag behind.

One way of explaining this phenomenon is that observers misjudge the time at which the static bar appears. We can only talk in this case about relative timing, we do not know exactly at what time we perceive each event. Also, in this case, as with many motion effects, getting the time wrong is the same as seeing the bars in the wrong position. We may have correctly perceived the relative time at which the flashed bar appeared, but misjudged the position of the moving bar.

The flash-lag effect has caused controversy over the years as it may tell us not only about motion detection but how we localise events in space and time. Hence the argument has raged amongst theories such as motion extrapolation (Nijhawan, 1994) and latency difference (Whitney & Murakami, 1998), both of which attempt to explain this effect. Although in the following modelling and experimental work I will not be addressing the question of the flash-lag effect

directly, we will be making use of and referring to some of the ideas that have emerged from this debate.

Motion extrapolation This is one of the earliest theories put forward to explain the flash-lag effect. It posits that we correct the position of a moving object by spatially shifting it in the direction of motion (Nijhawan, 1994, 2002). Shifting the moving object ahead allows us to correct for neural delays and see the object where it really is and not where it was when the light first hit the retina. This is an intuitively appealing argument, and the need for this correction can be seen when catching a ball. In order to place our hands in the right position we would need up-to-date information on the position of the ball. In order to be able to extrapolate, however, the motion needs to be predictable. Whitney and Murakami (1998) later went on to demonstrate in a simple experiment a counter-argument to motion extrapolation. In this case, a horizontal moving bar that suddenly reverses its direction is presented. At regular intervals along its trajectory the position at which a flashed bar appears aligned is measured. The flashed bar never appears to be in line with the moving one beyond the reversal point. The extrapolation model predicts extrapolation of the position of the moving object beyond the reversal point. See Fig. 1.5 below.

Fig. 1.5 The black line represents the trajectory of the moving bar and the dots are the points at which the flashed bar appears to be aligned. The red line is the predicted position of the perceptually aligned flashes as predicted by motion extrapolation. Reprinted with permission from (Whitney & Murakami, 1998) © *Nature Publishing Group*.

However, it has since been proposed that this lack of overshoot can be explained by backward masking, in that the new direction of motion interferes with our percept and causes us to ignore our previous perceived position (Nijhawan, 2002). Empirical evidence is still needed to establish this as a viable explanation. It is possible that extrapolation can apply in the case of predictable stimuli, but the question then arises whether the visual system distinguishes between predictable and unpredictable stimuli. It is an appealing concept, as in other areas of visual perception there are examples where the visual system appears to "fill in" over uniform areas (Gilchrist et al.; Komatsu et al., 2002). The extrapolation model has been generalised to instances of flash-lag in other domains, with extrapolation occurring for continuously presented attributes, versus temporary briefly flashed attributes of the visual scene (Sheth et al., 2000).

Latency difference An alternative theory posed is that the flash-lag effect is due to the differential delays involved in processing static and moving stimuli (Purushothaman et al., 1998; Whitney & Murakami, 1998). According to this model, the brain does not attempt to synchronise processes to ensure that relative timing is maintained, rather these differences are usually small enough to ignore and it is only in special cases such as the flash-lag illusion that we notice them. They argue that there is no specialised timing mechanism and that we perceive time only through the indirect measure of awareness of events. They claim that processing information about the location of a moving object is more rapid than for a static one. By the time we become aware of the position of a flashed bar we are already aware of a further position of the moving one, therefore we perceive these two positions as being simultaneous. This argument is supported by the fact that motion processing is known to be one of the fastest visual processes (Allik & Kreegipuu, 1998) and does appear to be processed in separate areas of the brain from static object properties.

However, there are certain caveats that need to be applied to this argument. It is not clear whether we can infer from the speed of detecting motion, that we are necessarily faster at detecting properties of moving objects such as location. One counter-argument is that even when the moving bar appears at the same time as the static bar (the flash-initiated paradigm), the latter appears to lag, when initially they should both be subject to the same processing delays (Eagleman & Sejnowski, 2000). Arnold et al. (2003) showed that the tilt after-effect does not behave according to the perceived relative position of moving and flashed gratings, but corresponded more with the stimulus as it would appear on the retina. They presented a stimulus consisting of a grating within a rotating annulus. In the centre of this annular grating a circular grating was presented briefly. The angle of perceived tilt of the central grating was measured as a function of the angle of tilt of the annulus grating at presentation. The tilt effect was consistent with the angle of tilt of the annulus as presented, rather than some more advanced position. This suggested that when position is considered as the relative orientations of gratings there is no

delay between registering the position of the static flashed grating versus the moving one.

The latency difference model predicts that aligned flashes will appear further along the same trajectory, but slightly behind in time. In the data above, however, the perceived flash positions do not reflect the trajectory around the reversal point. The authors suggest that this is caused by the added effect of temporal integration. It seems that latency difference by itself is not the whole story.

As is the case with extrapolation, the differential latency explanation can be extended to a general neural delay between continuously changing and sudden, briefly presented aspects of the visual scene. The criticisms of the extrapolation model also hold for this explanation. The idea of differential latency is linked to the idea of temporal facilitation, where a previously continuously changing object is more easily perceived than a sudden unpredictable attribute, even within the same feature domain (Bachmann & Poder, 2001). The differential latency model has also been further extended by allowing differential latency to fluctuate, with a probability density function approximated by a Gaussian function (Murakami, 2001).

Post-diction This model attempts to explain the smoothed shape of the perceived trajectory of the moving bar around the point of reversal as described above (Whitney & Murakami, 1998). In introducing this model (Rao et al., 2000) the useful distinction of event time (the time the event occurs), brain time (the time the brain has finished processing the information) and perceived time (when we become aware of the event) is made (also in (Johnston & Nishida, 2001)). As opposed to the latency difference model, the post-diction model differentiates between brain time and perceived time. It is proposed that before the final step of perceiving the bar there is an added process of optimal smoothing. This is a method often used in engineering to get more accurate estimates of measurements in the presence of noise. In order to estimate the position of a moving object at a given time *t*, measurements of position some

time (at times $t+\tau_1, t+\tau_2$ etc.) after the time t are taken into account. This will make our estimation of the position of a bar at time t more accurate. The added assumption is that the flash 'resets' this calculation, so that only positions after the flash are used in estimating the moving bar position and hence the spatially pushed forward estimation (Eagleman & Sejnowski, 2000). This account has been criticised as it implies some sort of positional leap from the correctly derived position of the moving bar from before the flash to the more advanced position after, and some sort of corresponding perceived velocity increase might be expected, which is not observed (Nijhawan, 2002).

Temporal averaging of position Krekelberg and Lappe (2000, 2001) concentrate on the specific calculation needed to retrieve the relative positions of two objects. They suggest that relative position is given by the difference between positions averaged over the time of the trajectory. This averaging period is around 600 ms. Briefly presented flashes or stroboscopic movement lag behind because of the temporal persistence of their signal causing a position difference between strobed and continuously moving objects, after averaging over time. This model references the effect of the temporal impulse response of the visual system, but it has been criticised on the basis of lack of physiological evidence for a sufficiently long averaging time (Nijhawan, 2002).

Sampling error Brenner and Smeets (2000) suggest that the flash-lag is due to the error in sampling a single position in time from the trajectory of the moving object. They suggest that after we receive the flash information, we set off a sampling process that takes the position of the moving flash. However, the sampling process itself takes some time and hence the sampled position of the moving bar corresponds to some later time. Again, this explanation can be generalised to flash-lag effects in different modalities. If we consider a sampling time for all continuous changes, then one can postulate that sampling time varies over different modalities. This explanation relies on the different nature of continuous and instantaneous stimuli. The sampling theory can be related to the theory of asynchronous feature binding in which continuously changing

features of an object are mis-bound with temporary features of an object (Cai & Schlag, 2001). These theories tend to remove the emphasis on low-level, bottom-up explanations for the flash-lag, interpreting it as a result of more cognitive processes. Attention based explanations would fit in this category (Baldo & Klein, 1995). However, although the flash-lag effect can certainly be attentionally manipulated, it has been shown that the lag is not simply a by-product of distraction by the flashed object (Khurana et al., 2000).

Broadly speaking, all explanations of the flash-lag effect fall into the category of temporal or spatial effects. However, we have mixed information about relative timing and spatial position, with the same marker for position and time, making it difficult to disentangle the two dimensions. We have highlighted some of the relevant ideas raised by the flash-lag debate. We are concerned with the interaction of motion and spatial position. If we are to consider whether this is a process based on feedback connections, then the evolution of neural response over time will have to be considered.

1.4.5 Shifts in perceived position of static objects caused by motion

More direct demonstrations of the effect of motion on spatial position have been shown. In the case of flash-lag, it is the relative position of the static object with respect to the moving object that is misperceived. In the following experiments the perceived location of two static objects relative to each other is tested and found to be affected by nearby motion. Although spatial localisation and timing are concepts that are easily confounded in the case of moving objects, we will now illustrate a few cases where in particular motion appears to distort the perceived location of objects. Misalignment between two static objects can also be found in the presence of motion. In this case the perceived misalignment cannot be said to arise as an artefact of having to make a comparison. The tasks involve alignment judgement tasks (e.g. vernier acuity) that subjects perform well. One such example of the effect of motion on spatial localisation was described by De Valois and De Valois (1991). In this case, static windows

limiting a moving Gabor pattern appeared to be shifted in the direction of the movement. It was previously found that these Gabor patch patterns could be aligned very accurately when just counterphase flickering. (See Fig 1.6)

Fig. 1.6 Stimulus for measuring vernier acuity in static moving Gabors (De Valois & De Valois, 1991). The top and bottom patches are flickered in counterphase, whilst the middle one is drifting to the right. A small misalignment is perceived between drifting and flickering patterns in the direction of motion.

Although the pattern is moving, it remains within a static window. Note a permanent hard edge reduces the illusion (Whitney, 2003). The misalignment is small (max. around 15 arc min) but consistent and has been replicated in further experiments, which will be described. Vernier acuity, which can be dissociated from the bias, was as good for moving as for flickering patterns.

The size of the mislocalisation increased with increasing eccentricity. The effect was greatest at 4-8 Hz temporal frequency, increasing with temporal frequency up to this point, before dropping off again for higher temporal frequencies. Hence, this effect does not increase linearly with speed. The shift was greatest at low spatial frequencies. Intriguingly, larger biases were found with patterns moving towards or away from the fovea than with those moving in a tangential

direction. Much of the following work in this thesis aims to investigate the possible mechanisms that could underlie such a motion induced position shift. Below we describe other attempts to investigate this phenomenon.

The possibility of finding a corresponding change in the activity in V1 was investigated using fMRI techniques (Whitney, 2003). Is it possible that representation in the primary visual cortex changes under the influence of motion, hence breaking the strict retinotopy that is normally found there? Indeed, it was found that Gabor patches arranged in the configuration shown below (Fig. 1.7) activated different areas of the cortex, depending on whether motion was inwards or outwards.

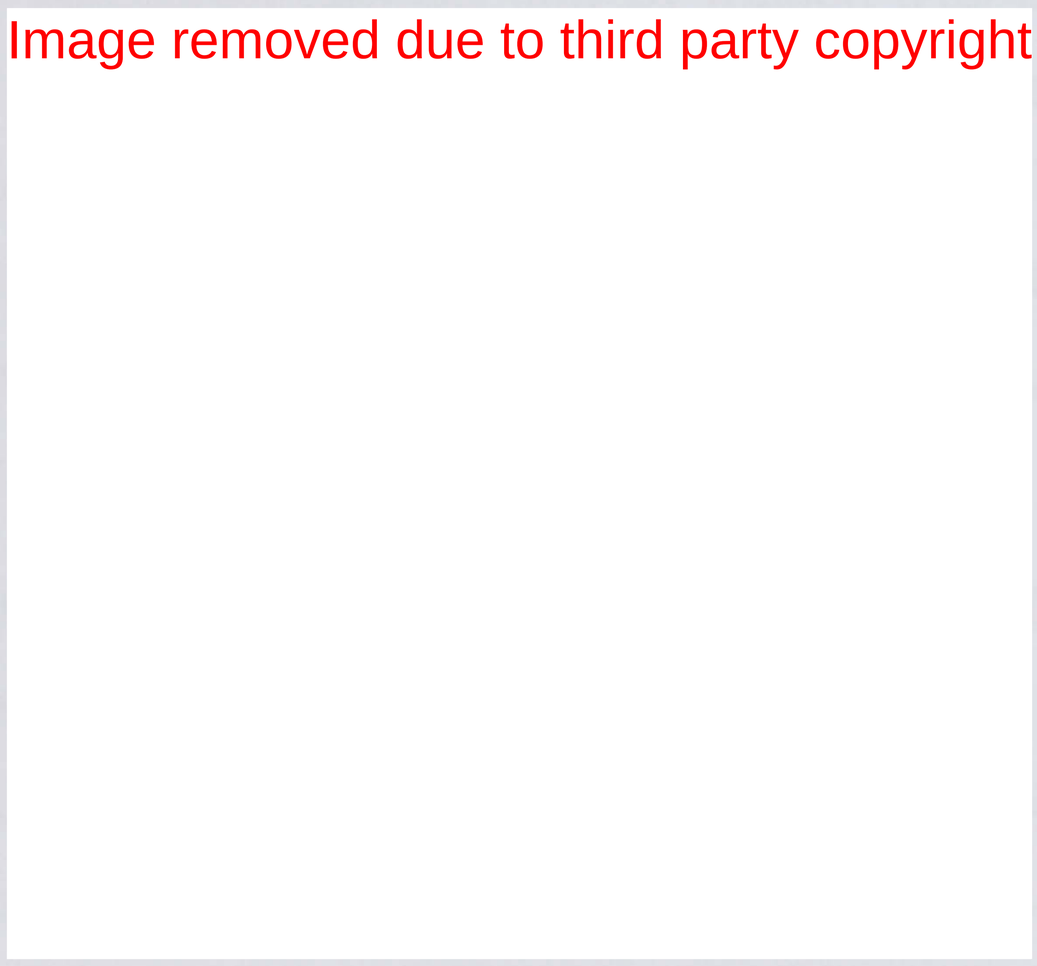


Fig. 1.7 Configuration of drifting Gabor patches (Whitney, 2003). Patches are presented either drifting towards the fixation point or away.

The location of the differences in activation was however counter-intuitive. They found that the inward motion activated a more peripheral area of the V1 cortical map than the outward moving patterns, which is opposite to the two percepts generated. They have demonstrated a triple dissociation between stimulus retinal location, the location of cortex activation and the apparent location of the stimulus. They also found the same difference for hard-edged pattern, where

there is no illusory shift, making it unclear if this difference in activation leads to the final percept.

This drifting Gabor patch stimulus and induced shift was also used to investigate whether the retinal or perceived position of the patches determine perceived contours in a pattern (Hayes, 2000). It was found that the contours were more salient when the perceived position of the patches were aligned, suggesting that the visual system dynamically extracts spatial position from the aggregate response of local computations.

The effect of translating motion on the static envelopes that contain it was also shown for random dot motion windowed within a static random dot background, with the edges defined only by motion (Ramachandran & Anstis, 1990). It was found that this type of positional shift was greatest at equiluminance and increased if the static dots were replaced by dynamic random noise. They also found that the shift could affect perceived size by presenting 'shrinking' and 'expanding' motion within circular envelopes, i.e. the shrinking circle appeared smaller than the expanding circle.

Further intriguing cases involve the effect of motion signals on position when there is no physical motion present. Nishida and Johnston (1999) presented a clockwise rotating windmill to the left of fixation that subjects adapted to over time. After adaptation the subject was presented with a test stimulus of two aligned windmills, either side of fixation. The left hand side windmill appeared to be tilted anticlockwise in the direction of the motion aftereffect. See Fig. 1.8.

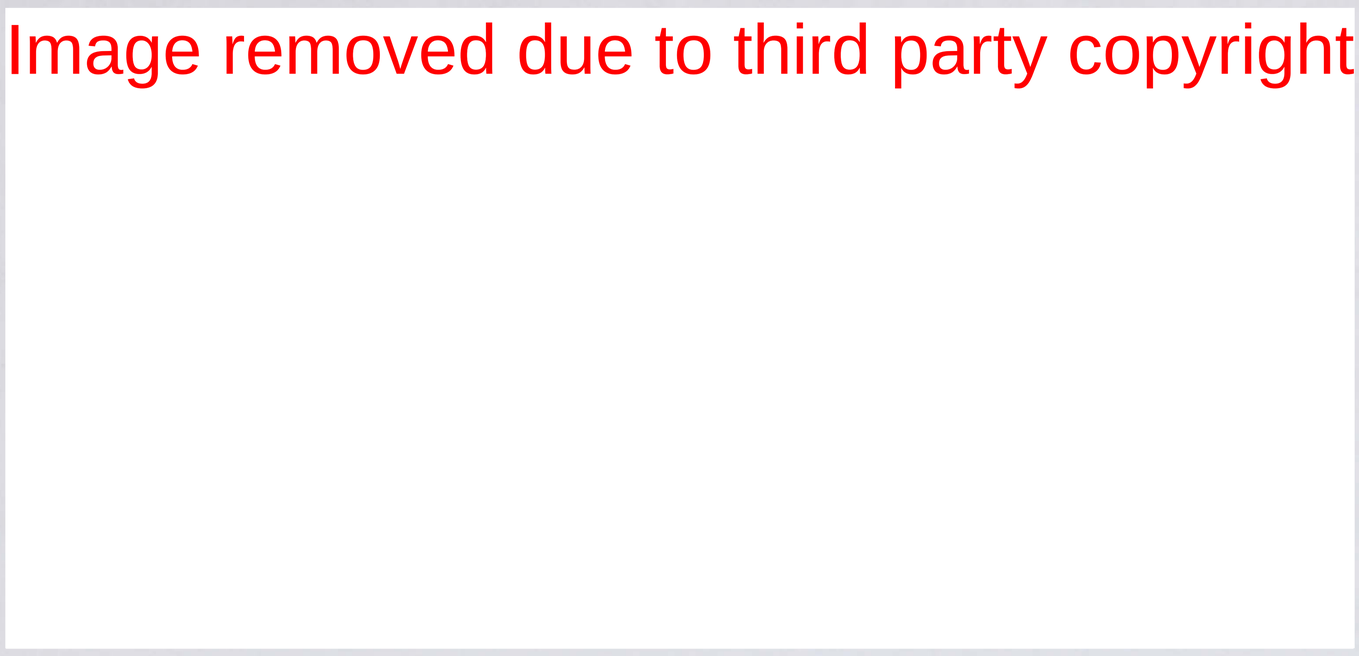


Fig. 1.8 Percept after adaptation to clockwise motion on the left (Nishida & Johnston, 1999), when the windmills presented are both physically vertical. As well as experiencing a sensation of anticlockwise rotation (the motion aftereffect) subjects perceive the windmill on the adapted side to be rotated away from the vertical.

They repeated the experiment, but this time after adaptation they nulled the MAE, by presenting a rotating windmill on the adapted side that rotated clockwise (in the opposite direction to the MAE) so that both windmills presented after adaptation appeared to be stationary. In this case the windmills appear aligned.

The shift in relative orientation of the left hand windmill was also measured, as a function of time after test presentation. It was found that the misalignment increased at the beginning of the test presentation (which the MAE does not), before the effect faded away with time (lasting longer than the MAE). They also found storage such that the size of the shift evolved from the time of test onset rather than the offset of the adapting pattern. It was pointed out that as both stimuli are static, misalignment cannot be due to processing latency differences. The authors raise the notion that position and motion may interact at several levels, suggesting possible recurrent input from area V5/MT to V1 or V2.

In a similar experiment, Snowden (1998) used two gratings moving vertically in opposite directions on either side of the fixation point as his adapting stimulus. This is comparable to the De Valois and De Valois experiment, but using the

MAE, rather than physical motion to induce a shift. After adaptation, test static grating patches presented either side of fixation appeared shifted in the opposite direction to the adapting motion (in the direction of the MAE) with respect to each other. As with the De Valois experiment he also found that the effect increases with speed up to a point, but drops off at high speeds.

He suggests that cells adapted to downward motion, for example, will fire less, causing a decreased signal for downward motion, but also a decreased signal for the particular location they represent, which is what might cause a disturbance in spatial localisation. So, rather than feedback, the explanation is suggested to lie in the dual task of V1 cells as position detectors and as parts of a motion detection system.

The parameter dependence of the position shift caused by the MAE was further investigated using the original Gabor patch stimuli (McGraw, Whitaker et al., 2002). Not only does physical motion not need to be present for motion induced mislocalisation, but it was reported by McGraw et al. that after adapting to motion and using test gratings with an orthogonal pattern, where there is no MAE present, one can still perceive a positional shift between an orthogonal grating on the adapted side versus one on the unadapted side. In contrast to Nishida and Johnston (1999) above, they managed to show a positional shift without the presence of perceived motion. However, it should be noted that the MAE can be experienced in the direction over which there is no luminance change in space, e.g. a sensation of “streaming” has been experienced when observing a blank field (Georgeson, 1976). McGraw et al. found a similar size of effect as was found with moving Gabor patterns and the same pattern with velocity. They also found that the way in which the adapting motion was generated made little difference; the shift was not tuned to spatial frequency or contrast of the adapting stimulus.

In the examples above, the position shift occurs to an object, which although static, has motion signals attached to it, either in the case of present motion or some form of the motion aftereffect. The next example involves briefly

presented static stimuli in the presence of continuous motion, as in the flash-lag demonstrations. However, the relative judgement is again between two static objects and not between the moving object and the static object.

This experiment also motivates much of the following work in this thesis as it raises important questions of spatial representation and motion calculation. The subject is presented with a central rotating black and white sinusoidal windmill. Two small horizontal bars are then flashed, one on each side of the windmill. The bars appear to be shifted with regards to each other in the direction of motion (Whitney & Cavanagh, 2000). See Fig. 1.9.

Fig. 1.9 The effect of motion on briefly presented static flashes. They appear shifted in the direction of motion (Whitney & Cavanagh, 2000).

This experiment was repeated with adjacent gratings moving in opposite directions and flashed bars flanking each side of the gratings. The bars were shifted from each other in the direction of each grating's motion, demonstrating the effect is not particular to rotational motion, but rather is induced by movement in general.

Several observations were made from this experiment. Listed below are a few of the key features:

- The misalignment remained constant as a function of the separation of flashes from stimulus, as long as motion remained central.

- The misalignment was present at the point of reversal of direction of the rotating windmill when there was no physical motion present.

- There was a noticeable effect of grating size (larger ones produced greater effect).

- The flashed bars are not perceived to be moving.

The motion dependence of the effect is shown by the relationship between grating size and misalignment size. Increasing grating size increases the motion signal size. Vernier acuity for two flashed lines of the size and separation used in the experiment is normally very good. When introducing the motion in-between the bars, acuity is not reduced, simply a bias appears in the direction of motion.

It appears that we can discount the account of flashes simply being caught up in a local motion field, as misalignment remains constant with flash separation and the flashes are not perceived to be moving. However, if motion is not central to the scene, the effect does not persist and drops off rapidly if the flashes are kept foveal and the motion moved further away. We know that in the De Valois and De Valois (1991) experiment, the positional shift increases with eccentricity, this leaves open the possibility that when motion is central, confounding distance from motion with eccentricity could affect the results. So, although the flashes do not appear to be moving is it possible that local motion could be affecting them in some other way? It is suggested by the authors that such long range effects might imply more higher level, cognitive binding cause for this perceived misalignment.

This example differs from the flash-lag results as the two flashes have exactly the same properties and are not themselves in motion, yet they are seen misaligned from each other. The misalignment is observed even at the point of

reversal of direction, when there is no physical motion present. The flashes are observed to be misaligned in the post-reversal direction. This could possibly be another manifestation of the flash-lag effect. In any case, it tells us a little about the timing of the influence of motion on position.

(Whitney & Cavanagh, 2000) mention that this shift may mean that the flash-lag effect is greater than previously measured. The question arises, in the case of a typical flash-lag stimulus with a moving object and nearby flash, does the flash undergo a spatial shift in the direction of the moving object's motion?

It was also found that the shift decreased with longer presentation times of flashes, i.e. permanent flashes were not found to be misaligned at all. A comparison may be drawn with drifting patterns presented in hard edged envelopes, which also do not appear shifted (Whitney, 2003). (Whitney & Cavanagh, 2000) also tested whether two flashes had to be straddling the central motion for a positional shift to occur. They measured the perceived position difference between a single flash presented next to upward motion and a flash presented at the same location, next to downward motion. To do this they presented a grating next to the flash position that reversed motion direction at some point. Two flashes were presented at the same point near this grating before and after reversal. A much reduced apparent misalignment between the two presentations was found than when the comparison was made between two flashes straddling opposing motion.

In the flash-drag case we see misalignment between two static flashes, which cannot be explained by latency delays, and at the same time these flashes are not perceived as moving and therefore it cannot be some kind of movement extrapolation. Does this indicate that this effect is completely independent to that of flash-lag and both effects require a different explanation? The authors suggest that the root of the problem may lie in a basic mechanism that subserves both spatial localisation and motion detection and it is the interdependence of these features that cause the effect rather than their separate processing pathways. The same cells in V1 that contribute to building

up a motion sensitive system must also contribute to our spatial localisation of objects and in some ways our timing of events. We know that the vast part of visual information from the retina passes through V1 and we know in addition that motion specific cells must also encode position in some way.

However, Whitney and Cavanagh (2000) also found that the misalignment occurs under binocular conditions, with motion presented in one eye and the flashes in another, implying a role for areas with binocularly driven neurons. The large spatial extent of the effect also might imply cells with large receptive fields and areas where more global aspects of the scene are processed. Hence, it is suggested that it is possible that re-entrant information from motion areas to V1 is the cause of the shift.

Further evidence for a more high-level effect of motion on position was presented by Watanabe et al. (2002). Two line-drawn diamonds translated horizontally in opposite directions, one above and one below the fixation cross, either behind an occluding surface with a narrow slit or without occluding surface. When the diamonds were in vertical alignment, two vertical bars were flashed, one in the centre of each diamond. In the case where the motion is seen through the slit, the percept is that of the object moving behind the slit even though the physical motion present is predominantly in the opposite direction to the motion of the perceived object. They found that the position shift occurred in the direction of the illusory motion where motion was only seen through a thin slit, and the shift was just as large as when the whole display was seen. The assumption would be that V1 responses would be tuned to the direction of the motion physically present, whereas it is more likely to be at higher level such as V5/MT that cell responses reflect the perceived direction of motion.

These experiments indicate that not only does motion seem to affect our representation of the relative position of moving and static objects, but it has an effect on where we see spatially concurrent and nearby briefly presented objects. It has been suggested that cells involved in detecting motion along the

visual pathway also contribute to relaying positional information available to us from the retina and it is this dual role that can cause such effects (Snowden, 1998). The possibility of recurrent mechanisms from higher cortical areas than V1 has also been suggested.

This is obviously an area where a modelling approach might prove useful, so that multiplexing and feedback loops can be represented quantitatively. "Quantitative modelling of the size of these effects is not yet possible" (Snowden, 1998), so if this were to be made possible, we would indeed have a useful tool.

1.4.6 Further effects of motion on the visual scene

The effects of motion on spatial position can be dramatically illustrated when they affect perceptual binding of aspects of the scene. Binding is simply the notion that all separate aspects of a scene need to at some point be bound together into one overall percept. It is not known if this process is carried out explicitly in the brain, but we assume that aspects of a scene that we perceive as happening at the same time are the ones we bind together. For example, if we see the car as being blue at the same time that it is moving right, then we see a blue car moving right. It is assumed that spatially and temporally co-occurring attributes are usually bound together. If motion disturbs where we see parts of the scene, then it can disturb how we attribute the characteristics of the moving objects.

One such example (Cai & Schlag, 2001) occurs in the following experiment. See Fig. 1.10.

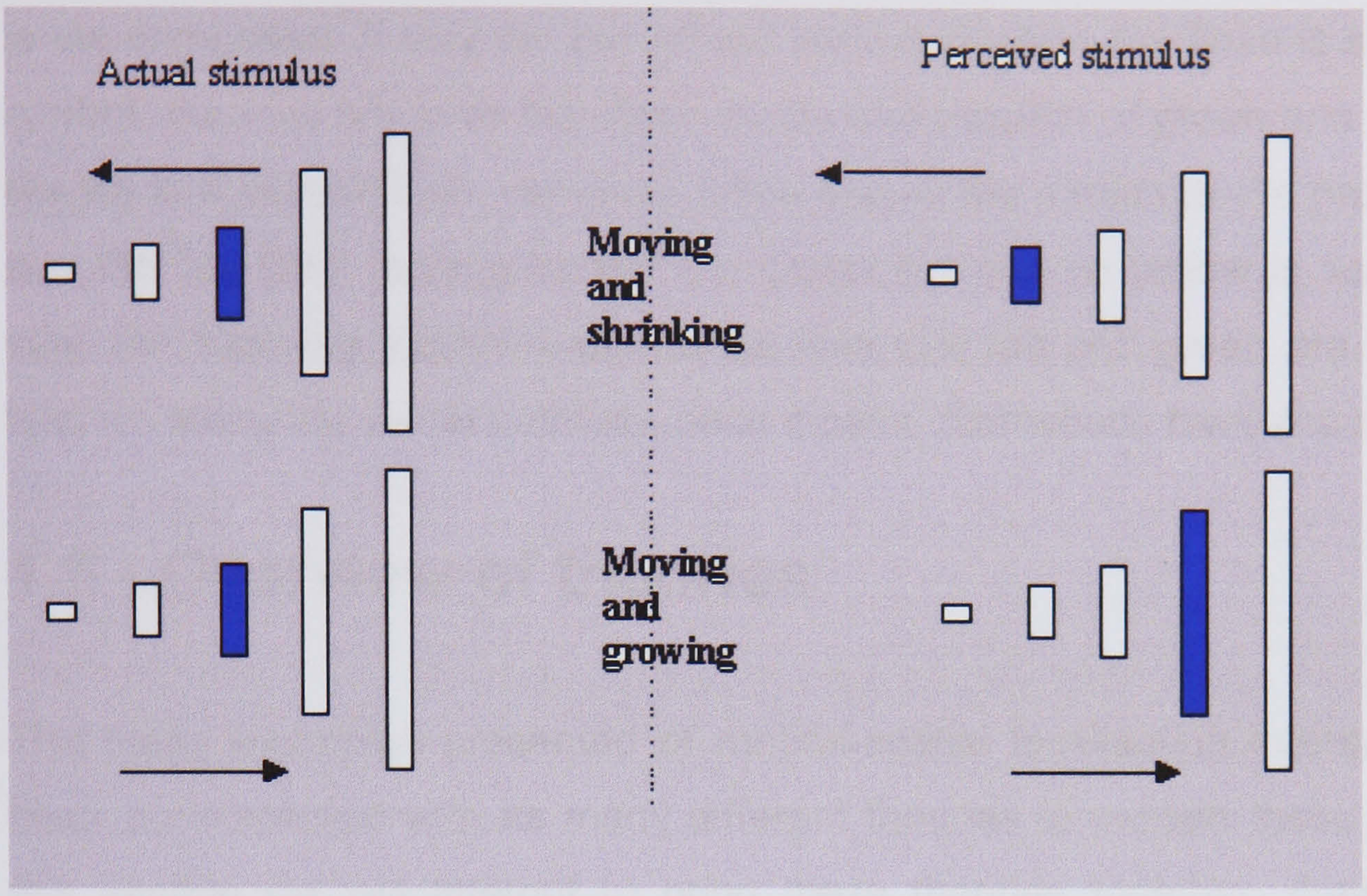


Fig. 1.10 Stimulus demonstrating asynchronous feature binding. On top is presented a bar moving to the right and growing in size in discrete steps, and on bottom a line moving to the left and shrinking. They are the same size and aligned in the middle, but at this position their colour turns blue. The blue bars appear misaligned and of different size as shown.

The white bars are presented as shown, one above the other, initially on opposite sides of the screen, on a grey background, moving in opposite directions, one shrinking, the other increasing in size, taking discrete positions and sizes. They reach the same size at the same location and time in the middle. The instant they are aligned in the middle and the same size they flash the colour blue. After the flash they turn back to white and they carry on shrinking/growing and moving across the screen. The percept is that of a smaller blue bar misaligned from a larger blue bar. The perceptual misalignment of the blue bars could indicate the forward shifting effect of motion on briefly presented static objects, but this does not explain the difference in perceived sizes of the bars. This intriguing example has many explanations involving all the theories listed above.

A related effect, involving colour, has also been demonstrated (Nijhawan, 1997). A continuous moving green bar is shown and at some point a red flash

is superimposed. If only the part of the screen showing the flash is shown as a control, then we perceive this flash as the combination of green and red, which results in a yellow flash. However, when shown the motion of the green bar we see the red flash lagging behind the green bar and no yellow is seen. In this way, the flash-lag illusion supports the idea that red and green are only fused later on along the visual pathway, once motion interactions have occurred.

1.5 - Overview of theories

The many examples presented of motion-spatial localisation interaction have been accompanied with as many different theories to explain these effects. In this section we will attempt to bring these theories together, draw parallels between them and see what further supporting evidence there is from psychophysics, physiology and experiments involving various other techniques.

1.5.1 Low level interactions in V1

The first level at which to look for the cause of interaction effects is the lowest cortical level, in V1. This is where the main part of initial motion processing occurs in humans and also where we still have the most accurate spatial representation of objects from the outside world. At the V1 level we can explain mislocalisation in the presence of motion as due to the dual role of cells (Snowden, 1998). We know that in V1 many cells are directionally selective to motion and that we may also need these to code positional information.

We might also consider the effect of cells on their immediate neighbours. In V1 cells are topologically arranged, corresponding to their receptive fields on the retina. Neighbouring V1 cells may have excitatory connections, causing the response from one cell to excite neighbouring cells and making them respond more quickly in turn. This would mean that receiving position information from a point in space may facilitate seeing information quickly from points surrounding it (Bachmann & Poder, 2001).

If motion and other scene properties are processed in completely separate streams, evolving separately to awareness as suggested by the modular view of processing, then any interaction would have to take place in "early" cortical areas such as V1. By looking at interactions at this level we may be able to explain some of the phenomena.

One possibility is that V1 cell responses are altered in the presence of motion through lateral connections between V1 neurons (Hirsch & Gilbert, 1991). Previous models of V1 have incorporated what is known about the nature of these lateral connection to show that more complicated responses can arise from the combined activity over time of V1 cells such as region segmentation, figure-ground segregation and contour enhancement (Li, 2001). Some models have also used horizontal connections to explain psychophysical evidence of orientation dependent apparent speed, linking V1 to motion effects (Seriès & Georges, 2002). However, it has also been suggested that lateral connections do not extend far enough to account for the full range over which cell behaviour can be modulated (Angelucci & Bullier). The extent of the connections reported is in particular less than the extent of the effects of motion described above.

1.5.2 Feedback loop

Some of the theories presented so far have suggested a role for V1 as part of a feedback loop involving higher level processes. This is motivated by the fact that areas of the brain specific for motion processing in the V5/MT region are so called "higher levels" of processing, whereas spatial localisation is at its most accurate at V1, the first stage of cortical visual processing. Interaction between the two processes may suggest recurrent information passing through V1, possible through the recurrent connections that have been shown to exist (Ffytche et al., 1995). The fact that positional shifts can be induced by the motion after effect and illusory motion, indicates the involvement of V5/MT as we would expect neural response in this area to correlate with the perceived direction of motion, which is necessary to determine the direction of the

positional bias. In V1 neural response would correlate more with the motion physically present in the visual scene.

Evidence of the more complicated role of V1 in vision processing was found through neurophysiological experiments conducted on monkeys (Lamme, 1995. Crucially, they found that the initial responses of neurons were characterised by the more traditional filter responses to local features usually assigned to V1 cells, while the later responses depended on contextual information and were possibly related to higher-level computations. In (Lee et al., 1998) the stimuli presented required complicated texture border discriminations and within-figure activation, which are normally thought to be higher level functions. The possible role of V1 in the later stages of processing, (although sounding like a contradiction in terms), may be important in investigating interaction effects.

Further evidence at the physiological level for an MT/V5 feedback loop to V1 has been provided by the method of temporarily inactivating V5/MT by cooling (Bullier et al., 2001; Hupé et al., 2001; Hupé et al., 1998). V1, V2 and V3 responses from monkeys to various stimuli where recorded with or without cooling. The responses to a line moving over a background were recorded. The biggest difference between conditions (cooling vs no cooling) was found for low salience stimuli, implying cortical feedback improves discrimination between figure and ground (Hupé et al., 1998). It was also found that feedback from MT had a very early effect on neural response (within 10 ms).

(Foxe & Simpson, 2002) used ERP (event related potential) techniques to determine the flow of activity in early visual processing. Typically upon presentation if visual stimuli there is an early negative response over the scalp, localized around a single area of the scalp around the occipital cortex, with a peak of latency 50-60 ms. Foxe & Simpson (2002) found that after a mean onset latency of activity over the occipital cortex of 56 ms, the dorsolateral frontal cortex is active by just 80 ms. The early response of the dorsolateral cortex, given that activity in visual sensory areas typically continues for 100-400

ms prior to motor output, led the authors to hypothesise that there was time for repeated iterations of feed forward and feedback loops.

Pascual-Leone and Walsh (2001) used a TMS (trans-cranial magnetic stimulation) experiment to investigate such feedback connections. TMS can either induce a percept of phosphenes (suprathreshold) when applied to visual areas of the cortex or temporarily "knock out" a visual area (subthreshold). It has been shown that applying a suprathreshold TMS pulse to V5/MT causes the perception of moving phosphenes. The authors found that by applying a subthreshold TMS pulse to V1 *after* a suprathreshold a pulse has been applied to V5/MT, the perception of moving phosphenes is severely disrupted. There was no effect if the V1 pulse was applied earlier than the V5/MT pulse. This seems to be indicate that the activity from V5/MT coding for the moving phosphenes passes back through V1, where it is disrupted by the subthreshold pulse applied at the later time to V1. This has led to the suggestion that V1 may act as an "active blackboard" for integrating the results of calculations from different parts of the visual scene (Bullier, 2001a).

1.5.3 Brain time versus event time

When investigating the neural mechanisms underlying the interaction of two visual computations and how the two might be combined, we need to consider theories on how information across different modalities of the brain is integrated. In particular this involves the question of timing – how do separate parts of the percept combine over the temporal integration time of the visual system?

Past work has attempted to ascertain the effect of processing times in separate visual modalities. Moutoussis and Zeki (1997) investigated the relative timing of motion and colour perception. It was found that when alternating upward moving green dots with downward moving red dots, in order for the colour change to appear in phase with motion change, the motion direction change had to occur earlier. This was seen as evidence of a longer processing time for

motion than colour, directly leading to an asynchrony in awareness of each attribute. These results were reproduced (Arnold et al., 2001) using an entirely different method, reliant upon colour contingent adaptation. They found a maximal effect of adaptation if motion changed earlier than colour.

However, it has been argued that the Moutoussis and Zeki effect could depend on the rate of changes of colour and motion. By making the colour change a gradual continuous one and the change in position an instantaneous change between higher and lower positions, the opposite effect was demonstrated. The colour change needed to be presented earlier than the position change for the changes to appear in phase (Nishida & Johnston, 2002). Nishida and Johnston argue against this passive latency difference explanation of timing effects proposed by Moutoussis and Zeki (1997). They say that if we consider movement to be a direct change in position, then change in the movement direction is a more indirect, removed change in position, whereas change in colour is a direct change. They argue that it is this difference that causes the mistiming. It is suggested that it is the more indirect change that lags behind.

These demonstrations are part of the debate about whether our perception reflects brain time or event time. Is the relative perception of the timing of events locked into the relative timing of the neural response as sensations evolve in spatially separate parts of the brain? On the other hand is it possible that these separate sensations are combined in such a way that attempts to calculate their original relationship to each other, i.e. the brain attempts to correct for relative delays?

We have seen that the latency difference explanation of the flash-lag effect may be extended to the Moutoussis and Zeki (1997) demonstration, and also to the Cai and Schlag (2001) example. However, how do we generalise which aspects of the scene are processed more quickly? In the flash-lag case it was suggested that the position of the moving object is processed more quickly then that of the static one. In the Moutoussis and Zeki demonstration, the colour change is processed more quickly than the motion change, yet in the Cai

demonstration, the position of the moving object is processed more quickly than the colour of the flash.

Further, Arnold and Clifford (2002) showed that the delay between motion and colour could be manipulated by altering the relative direction of motion, before and after direction change, with increased delay for orthogonally opposite motion directions. Again, this demonstrates that percepts can be altered by manipulating relative delay.

Overall (as opposed to relative) effects of delay are clearly evident in the visual system. A clear example is that light intensity affects how quickly we are aware of objects. In cricket "bad light stops play" as at low light intensities we do not have up-to-date information on the position of the moving ball. In this case we are subject to neural delays that we are not able to correct for. Rapidity of neural response does correspond with the speed of awareness. The question is whether or how different processing speeds in different cortical areas affect our percept. Does the brain attempt to correct for either of the two: delay from reality or delay between visual modules?

Another source of timing error instead of the relative delay between visual modules could be the delay caused by the need to sample the position of moving objects, or even sample a state of a continuously changing feature (Arnold et al., 2003; Brenner & Smeets, 2000). If we take the case of continuous versus temporary objects, then there seems no reason that there should be a lag in the Moutoussis and Zeki case as these are both sudden changes.

Latency differences between visual modules cannot explain the mislocalisation effects we have seen as the misalignment is between static objects that do not change their physical position over time. However, the relative evolution over time of the motion percept and localisation of objects is interesting to us when attempting to explain these phenomena. For instance, we would like to know how motion affects the percept of position over the time of the impulse

response to a static flash since, "... cognitive processes cannot be understood without their temporal dynamics" (Poppel, 1997).

The time course of perceptual events is not just a question for vision and the same questions can be posed about temporal processing within other sensory areas. An interesting question is whether we use the same mechanism to order events in time across all the senses. Temporal perception also leads on to the question of binding – how we glue together aspects of the scene and decide they happened at the same time. This can be examined cross-modally and also within vision it is easy enough to find examples of the mis-binding of different visual attributes.

1.5.4 Discussing consciousness

Although we are concerned with low level effects, hoping to find a clear link between physiology and overall perception, we cannot avoid the question of consciousness. In order to report a percept we must be aware of it. It is where, how and when this awareness is reached that is hotly debated. Many of the theories discussed so far have implications when discussing consciousness.

The latency difference argument is appealing as it involves no higher level decision making process for comparing event times and is tied in with the modular theory of the brain. Daniel Dennet's paper, "Are we explaining consciousness yet?" (Dennett, 2001) also proposes this type of model. It maintains that consciousness should not be thought of as a separate step, but rather is the implicit effect of a certain level of activity of part of a brain. Timing of events would depend upon which events were creating enough activity at a certain time to become a conscious percept. Zeki also proposes this approach with his theory of "microconsciousness" (Zeki & Bartels, 1999). Following on from his modular approach and his experimental results he proposes that awareness of each percept evolves independently.

The argument is that any kind of further binding mechanism that might compare times or bind events together only removes the problem of explaining where consciousness is to a further detached level, where the question remains unsolved. In other words, it is the state of a brain with different active percepts that is imprinted and when asked which ones co-existed we simply decide which state was present. This point of view leaves open for debate what level of activity is necessary for awareness. Without any binding process it is not clear how the brain makes a difference between seeing two different shapes as separate or as a whole.

On the other hand, it is hard to envisage a special higher level area where consciousness resides and percepts are combined, especially as this tends to lead us back to the Cartesian theatre and the question of the homunculus. The possibility of consciousness emerging from feedback loops is an interesting spin on the question. Indeed, it has been suggested that, for example, feedback from area V5/MT to V1 is necessary to perceive motion (Pascual-Leone & Walsh, 2001). In this way there are no conscious versus unconscious pathways or modules of the brain; it is the combination and timing of the combination of areas in a certain way that is necessary for a conscious percept.

1.6 - Questions posed

We have uncovered many interesting examples that give us some insight in to the functioning of the visual system. It seems that to tackle the question of motion processing as part of an integrated system we cannot ignore the spatial localisation and the time course of visual processing. We have seen that perceiving motion can cause us to misjudge the relative positions of both moving and static stimuli. In the case of moving stimuli this can also be interpreted as the misjudgement of the timing of an event. The breadth of examples makes it difficult to find a theory to which there are no exceptions. It is also possible that different theories may apply in different cases. By probing

the causes of visual motion illusions in which we fail to represent the real world reliably we may re-address the old question of defining the mechanisms of motion processing. New models may not only explain apparent interactions over processing areas, but also may provide a more accurate computational representation of the human motion detection pathway. Through tackling motion processing we may then be able to extend theories of interaction to processing in other areas of brain.

Chapter 2- Empirical investigation of the motion induced spatial shift

As described previously (see Chapter 1), moving visual pattern can influence the perceived position of outlying, briefly flashed objects. In this chapter there follows an experimental investigation into the effect of spatially translating discrete objects on the perceived position of nearby briefly presented static objects. A moving pattern such as a drifting grating or a rotating windmill has motion associated with it that always occupies the same space and is present constantly over time. In the following experiment the motion present in the visual display is generated by a rotating bar. This will allow us to localise the effect of motion in time and space. First, there follows a brief recap of the existing literature that led to these specific experiments and the questions that initiated the investigation will be set out. Then the empirical work will be described in detail and finally discussed with possible implications and modelling questions. (The work described in this chapter is published in (Durant & Johnston, 2004)).

2.1 - Questions posed from previous work

Functional mapping of cortical areas has led to a modular, distributed view of visual processing in humans, each module with its own function and temporal characteristics (Zeki, 1978; Zeki & Bartels, 1999). However, this view provides little insight into how modules interact with each other to form a temporally coherent percept (Johnston & Nishida, 2001; van de Grind, 2002). The

perception of movement usually (but not always) coincides with a change in perceived position, implying coordinated activity in V5/MT and V1, but it is not clear how these two areas interact. In the light of growing evidence that motion, temporal and spatial position mechanisms do not operate in isolation (see Chapter 1) it is worth further exploring if the larger motion selective cells in V5/MT contribute to the motion shift. In Chapter 1 studies were described that suggest that the spatial shift is a consequence of a feedback pathway from V5/MT to V1. It would be interesting to see if any further indications of this feedback loop could be found experimentally.

Previously it was shown that the motion of a rotating or translating pattern can cause a spatial shift in the position of briefly presented, static, objects located some distance from the motion (Whitney, 2002; Whitney & Cavanagh, 2000). The question posed in the following experiments was whether a single moving object, generating locally changing motion signals could cause a mislocalisation of nearby flashes. If this shift could be generated we could then consider the effect of motion signal that changes over time on perceived spatial position. This would be a way to further explore the temporal dynamics of the influence of motion on spatial position.

It was also important to examine the effect of distance between a moving stimulus and the test bars with their eccentricity kept constant, to see if there is any decrease in effect size if the flashes are further from the motion. This would indicate a spatio-temporally localised effect of motion and suggests the spatial shift is mediated by low-level mechanisms rather than higher level/grouping mechanisms as has been suggested previously (Watanabe et al., 2002; Whitney & Cavanagh, 2000).

2.2 - Experiment 1: Varying the presentation time of the flashes

The first experiment determined the relative position (and corresponding relative time) over which a moving bar influenced the perceived positions of static flashes.

2.2.1 Methods

Stimuli (Fig. 2.1) were presented on a high resolution CRT monitor (800 × 600 pixels, 80 Hz refresh, SONY GDM-F520) controlled by a VSG graphics board (VSG2/3F www.crsltd.com) programmed in Matlab (www.mathworks.com) on a PC (www.dell.com). In all experiments subjects were seated 92 cm from the visual display. Subjects had normal or corrected-to-normal visual acuity. All parts of the stimuli were black (0 cd/m^2) and were presented on a white (53 cd/m^2) background.

The experiment took place in a dim ambient light. The rotating anti-aliased bar subtended 162 × 12 arc minutes of visual angle. The flashes were 11 arc min × 4 arc min and separated from the bar by 24 arc min. Subjects were asked to fixate on the middle of the rotating bar. Each trial consisted of the clockwise or anticlockwise rotation (40 rpm) of the bar for 2.5 seconds (for 1.7 rotations), during which time the two flashes were presented horizontally either side of the bar - three times for one frame (13 ms) every half a rotation. From trial to trial the flashes were vertically offset from one another about the horizontal (the offset varied between 10 arc minutes separation in the direction of motion to 21 arc minutes against the direction of motion). Subjects judged which flash appeared vertically higher and responded left or right by pressing a button. The number of responses (out of 20) against the direction of motion were recorded for nine values of vertical offset. This data was used to establish the point of subjective equality using probit analysis (Finney, 1971). The Method of Probits

involves fitting the integral of a normal distribution to the psychometric function. The 50% point on this sigmoid curve (Fig. 2.2) gives the point of subjective alignment. The slope of the curve, or equivalently the standard deviation of the underlying error function, provides a measure of the discrimination threshold.

Clockwise and anticlockwise presentations were interleaved randomly and since there was no noticeable effect of direction of rotation *per se*, the results were combined together into 'with direction of motion' and 'against direction of motion'. The angle between the rotating bar and the vertical at the time of the flash was varied across blocks of trials to measure apparent flash alignment for 15 moving bar positions (every 12° from the vertical) at flash onset.

Flash-lag for each subject was determined by presenting half a rotation of the bar, for on average 0.75 s and systematically varying the angle of the rotating bar at which a single instance of paired flashes were presented at the horizontal. Starting points and ending points were randomly jittered independently between 20° rotation about the vertical. Subjects were asked if the rotating bar was spatially ahead of the flashes or behind the flashes at the time of presentation. Four estimates (each based on 70 trials) of the 50% point on the psychometric function were averaged for each subject to determine the subjective temporal coincidence of flash and bar along with associated standard errors. The author SD, and two naïve subjects participated.

Experiment 1 was repeated, using the same method for 11 naïve subjects, with perceived misalignment measured for four relative positions of the bar to the flashes (0°, 60°, 90°, 150° past the vertical).

2.2.2 Results

At most positions of the rotating bar, two physically aligned flashes, presented horizontally on either side of the bar, appeared to be misaligned in the direction of bar motion (Fig. 2.1), but the magnitude of the effect varied significantly over different positions of the bar.

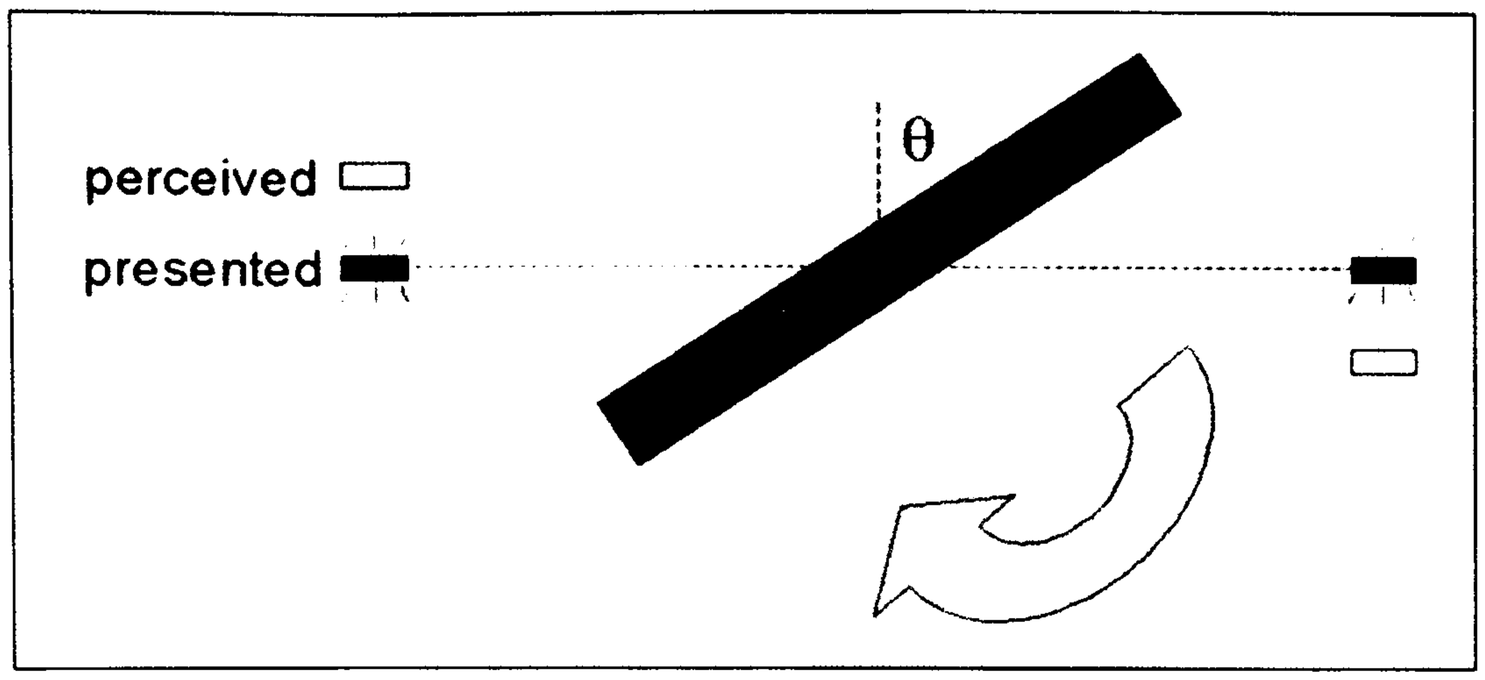


Fig. 2.1 The stimulus is a rotating bar (anticlockwise or clockwise) with flanking flashed bars. When aligned flashes are presented at a given value of θ, they are perceived to be misaligned, as illustrated, in the direction of the motion.

A typical psychometric function for one subject, SD is shown in Fig. 2.2a. We observe that there is a significant perceived misalignment in the direction of motion (11 min arc, SE=0.7 min arc). The standard error of the approximated subjective point of alignment is calculated along with the probit fit (Finney, 1971). Fig. 2.2b shows the plots of perceived misalignment in the direction of motion against the angle of the rotating bar at which the flash was presented for three subjects. We find that the size of the effect varies significantly with the point in the trajectory of the moving bar at which the flashes were presented.

Importantly, at no point is a comparable misalignment observed in the opposite direction to motion. Only at one point does one subject see a significant misalignment of a 1.8 arc min against the direction of motion (subject SD at 120°), whereas perceived misalignment in the direction of motion peaks for the same subject at 14.7 arc min. The fact that the perceived misalignment is almost always in the direction of motion means that the shift is not attributable to a simple spatial tilt illusion alone (Gibson & Radner, 1937; Wenderoth & Johnstone, 1988). A typical tilt illusion would result in equal and oppositely signed spatial shifts for opposite relative orientations of the bar with respect to the horizontally oriented flashes (Arnold et al., 2003).

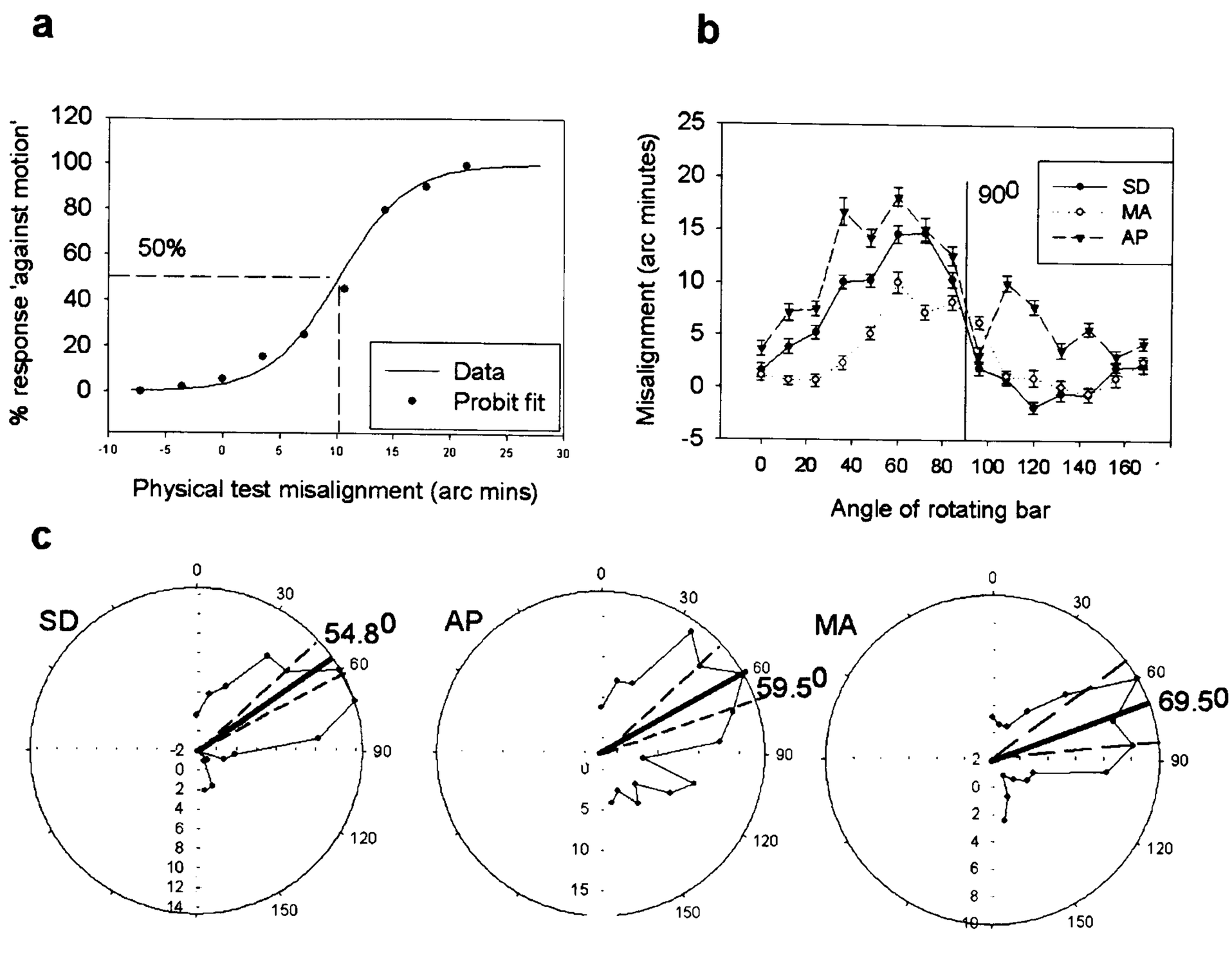


Fig. 2.2 Results for the first experiment, subjects SD (author), MA and AP (naïve). (a) Psychometric curve for subject SD. Flashes were presented when the rotating bar was six degrees before the horizontal. Responses 'right' vs 'left' higher are combined into 'with' or 'against' the direction of motion (percentage 'against' plotted). For the physical flash misalignment values, positive values represent flashes physically shifted against the direction of motion (nulling the effect). By checking the 50% point we see that subject SD perceived the flashes to be aligned when they were misaligned by about 11 min arcs all together against the direction of motion. (b) Plots of perceived misalignments for each subject against the angle from the vertical of the rotating bar at the time of the flash. A negative value corresponds to a perceived misalignment against the direction of motion. Error bars ±1 S.E. were found from the probit fit. (c) Data from (b) plotted on polar axes. Perceived misalignment is shown on the radial axis (arc min) and the angle at flash presentation on the angular axis. The zero circle indicates no misalignment and negative values indicate a misalignment against the direction of motion.

$$\text{Phase} = \tan^{-1}\left(\frac{\sum_{i=1}^{15} M_i \sin 2\theta_i}{\sum_{i=1}^{15} M_i \cos 2\theta_i}\right), \quad \text{Magnitude} = \frac{2}{15}\sum_{i=1}^{15}\sqrt{(M_i \sin 2\theta_i)^2 + (M_i \cos 2\theta_i)^2} + \bar{M}_i .$$

The phase was divided by 2 to find the peak angle. The error on this angle is calculated by drawing 1000 bootstrap samples from the normal distribution given by each psychometric function at each observed point and recalculating phase each time. Corresponding times between flash presentation and the bar reaching the horizontal are 147 ms, 85 ms and 126 ms respectively.

In order to determine the angle along the trajectory of the moving bar at which the presentation of the flashes results in a peak misalignment, we calculated the phase of the second harmonic of the data for each subject (Fig. 2.2c). We used the second harmonic as the data necessarily repeats every 180° with each rotation of the bar. Effectively we are fitting a sin(2θ) function to the data. We found that for each subject the peak misalignment (10-18 arc min) occurred when the rotating bar was about 30° before the horizontal or, equivalently, about 120 ms before the rotating bar reaches the position physically closest to the flashes (SD 147 ms; MA 85 ms; AP 126 ms). This temporal window lies within the temporal range of the flash-lag effect (Eagleman & Sejnowski, 2000; Krekelberg & Lappe, 2000; Nijhawan, 1994; Whitney & Murakami, 1998). This suggests that the size of the positional shift could be related to the *perceived* position of the rotating bar at the time of the flashes. However, the average flash-lag effect in this case, measured explicitly for the three subjects was only 24 ms (SD 41.1 ms; MA 13.9 ms, AP 17.6 ms).

The error on the fitted peak angles (Fig. 2.4) was found by bootstrapping as the time constraint meant we had only one measurement of perceived misalignment for each angle, but it was important to gain an idea of how much this peak angle could vary. As each measure of subjective alignment is taken from the fitting of a cumulative Gaussian function to the subject responses, we can take each perceived alignment to be the mean of a normal distribution of measurements with the standard deviation found from the slope of the cumulative Gaussian, which is one of the parameters returned by probit analysis. For each angle then we could sample this distribution 1000 times. For each sample we can calculate the peak angle. This gives us a distribution of a

1000 peak angles and the standard deviation of this distribution gives the standard error of the peak misalignment.

To establish the reliability of the spatial shift effect, we repeated the experiment over a group of 11 naïve subjects for four positions of the moving bar (0°, 60°, 90° and 150° from the vertical). In Fig. 2.3a we can see that a significant misalignment occurs at all the angles except 150°, (0°: $t_{(10)}$= 3.46, $p<0.05$, 60°: $t_{(10)}$ = 3.38, $p<0.05$; 90°: $t_{(10)}$= 3.73, $p<0.05$; 150°: $t_{(10)}$= -0.17, n.s.;), which is where we might expect the least effect from the first part of Experiment 1. The significant misalignment at 90° indicates that despite the relatively small size of the induced spatial shift (10 arc min), position is still disrupted when the bar is physically aligned with the flashes. We fitted a sin(2θ) curve to visualise how these four points might lie on a distribution over all angles (Fig. 2.3b). We can see that the data fit the shape of the distribution we found for the first three subjects in Experiment 1 and the peak of the sine curve lies at 19° (74 ms) before the horizontal, reinforcing the estimate of the time lag.

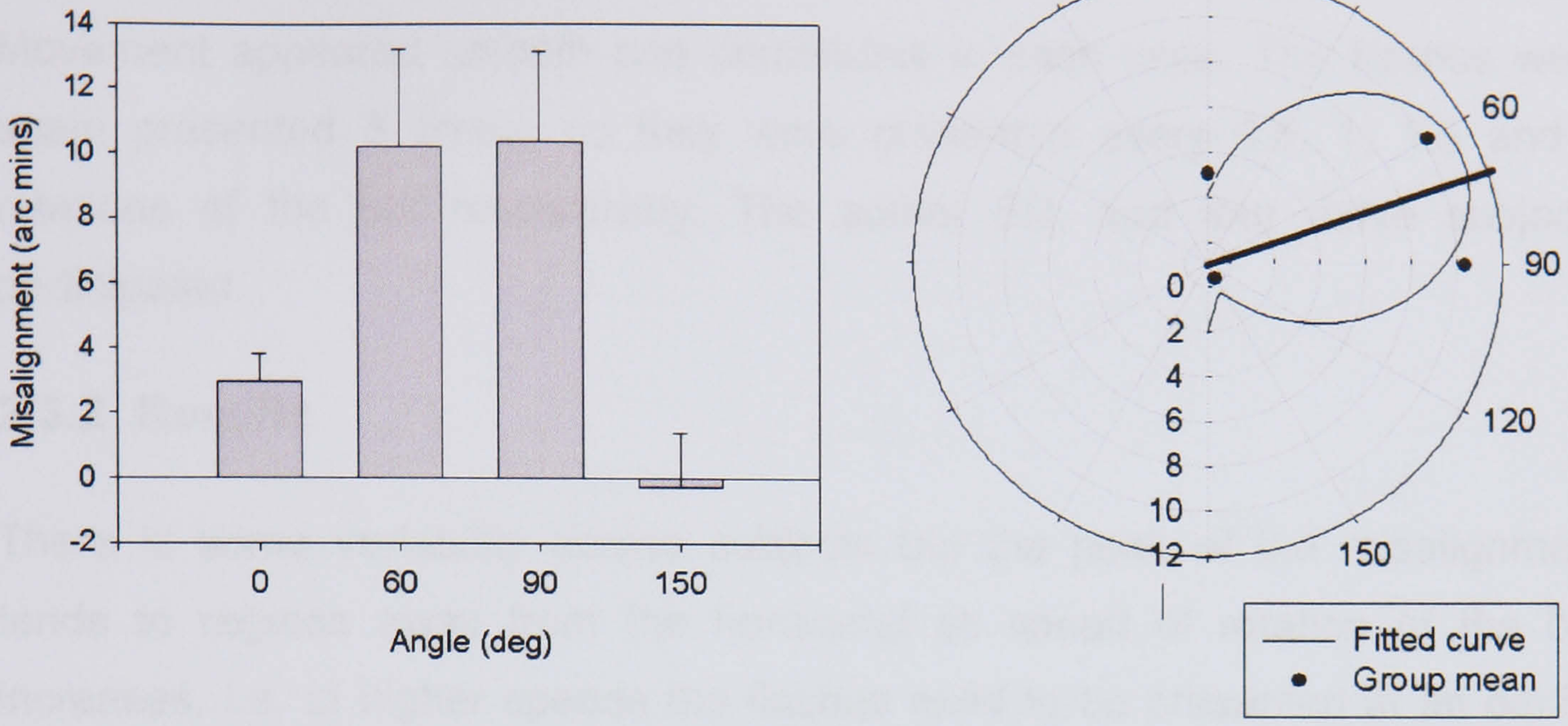


Fig. 2.3 Results of measuring perceived misalignment for 11 naïve observers. (a) We observe a significant difference over the four conditions. Error bars ± 1 S.E. There is significant misalignment in the direction of motion at 90° (when the flashes are presented when the bar is horizontal). There is no significant misalignment against the direction of motion at 150°. (b) The data from figure (a) plotted on a polar plot to illustrate how it relates to the distribution discovered in Experiment 1 (fitted with $a[b+\cos(2(\theta+p))]$), with a peak of 19° before the horizontal. Misalignment is presented on the radial axis (arc min); angle at flash presentation on the angular axis.

2.3 - Experiment 2: Varying the speed of rotation

In this experiment we tested whether it was the physical location of the bar at the time of the flashes or the timing of the flashes along the trajectory of the rotating bar that was crucial in determining the size of the perceived misalignment.

2.3.1 Methods

Using the same methods we repeated Experiment 1 (10 responses per test level), with the original speed of the rotating bar (40 rpm), and the original

speed × 2 (80 rpm), × 3 (120 rpm) and × 4 (160 rpm). The frame rate remained the same and the bar rotated 3°, 6°, 9° and 12° on each frame respectively. Movement appeared smooth and continuous in each case. The flashes were again presented 3 times, so they were presented every 0.5, 1, 1.5 and 2 rotations of the bar respectively. The author SD, and four naïve subjects participated.

2.3.2 Results

There is some variability across subjects but the peak of the misalignment tends to regress away from the horizontal as speed of rotation of the bar increases, i.e. at higher speeds the flashes need to be presented at an earlier point of the trajectory of the rotating bar to achieve the same size of effect (Fig. 2.4). If we average over the difference between the angle of greatest effect and the horizontal (Fig. 2.5a), we see an increase in peak angular difference with speed of rotation. We found a significant main effect of speed on the angular difference, $F_{3,12} = 12.84$, $p < 0.05$. If we plot the time between bar and flash position (rather than the rotation angle) that delivers the greatest spatial shift against the speed of the bar, there is no systematic change with bar speed for the five subjects, $F_{3,12} = 1.075$, $p = 0.396$ (Fig. 2.5b). The temporal difference averaged across subjects remains constant over all four speeds at a value of 62 ms. Following Whitney and Cavanagh (2000) we found no overall change in the magnitude of the peak perceived misalignment as a function of speed, $F_{3,12} = 0.583$, $p = 0.637$ (Fig. 2.5c).

We have found that it is the relative motion over a fixed time after the flashes are presented that is crucial, not the spatial position at the time of the flash. Note the flash-lag effect behaves in a similar way, increasing in spatial extent with speed according to a constant time rule (Nijhawan, 1994). This re-enforces the conclusion that there is no tilt illusion present as it has been shown that the angle of peak effect for the tilt illusion between a moving and flashed stimulus is not affected by the speed of rotation (Arnold et al., 2003).

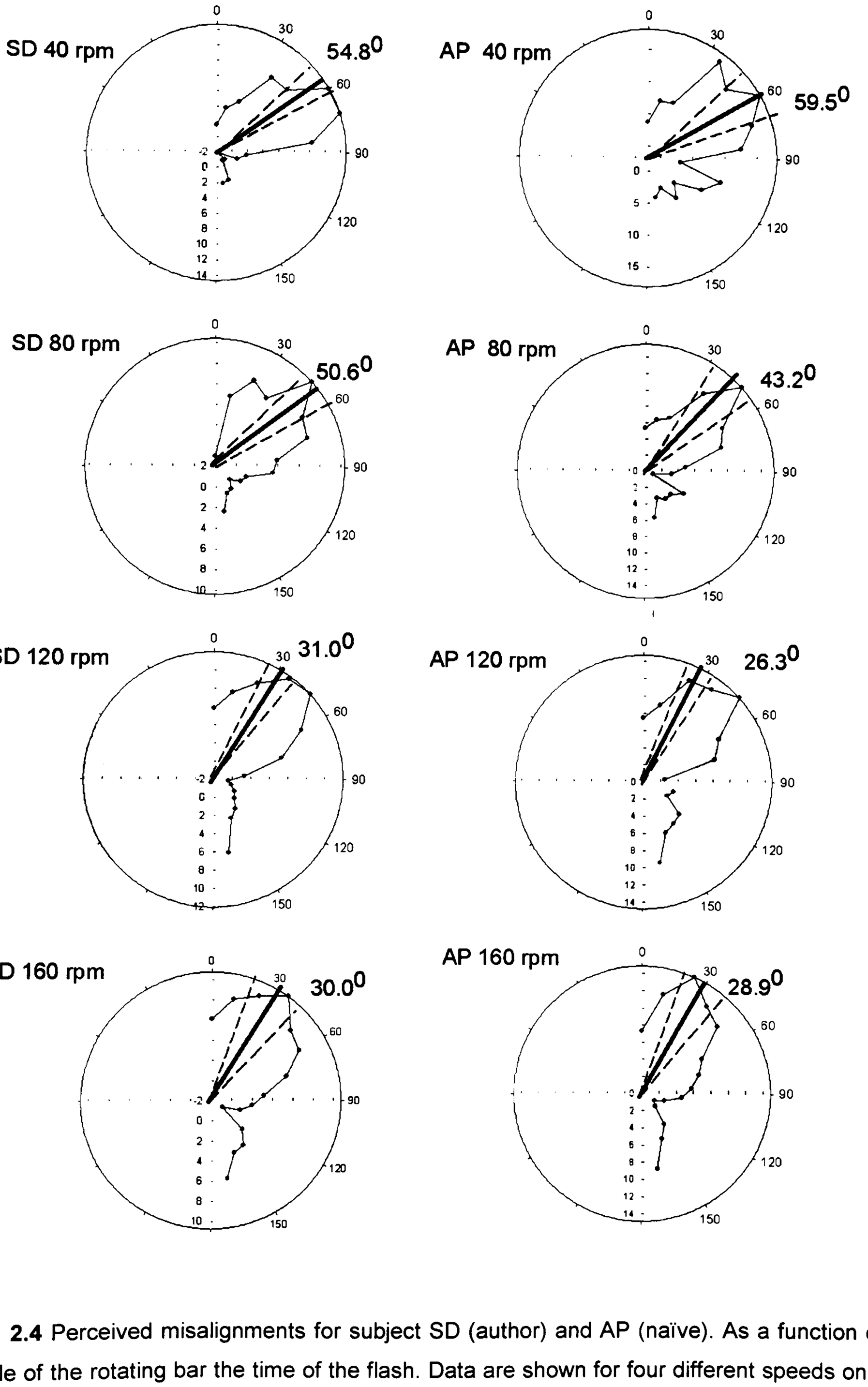


Fig. 2.4 Perceived misalignments for subject SD (author) and AP (naïve). As a function of the angle of the rotating bar the time of the flash. Data are shown for four different speeds on polar plots (misalignment on radial axis (arc min), angle of flash presentation on angular axis), with

associated peak angles expressed in degrees from the vertical. With increasing speed, the peak angles move further away from the horizontal. S.E. error bars calculated by bootstrapping as before.

The use of a spatially localised moving stimulus has allowed us to measure a spatio-temporal window over which movement can have influence on the bar. The critical determinant is motion introduced near the test bar locations after the flash. This is consistent with Whitney and Cavanagh's (2000) finding that flashes presented at the time of a change in direction of rotation go with the following motion.

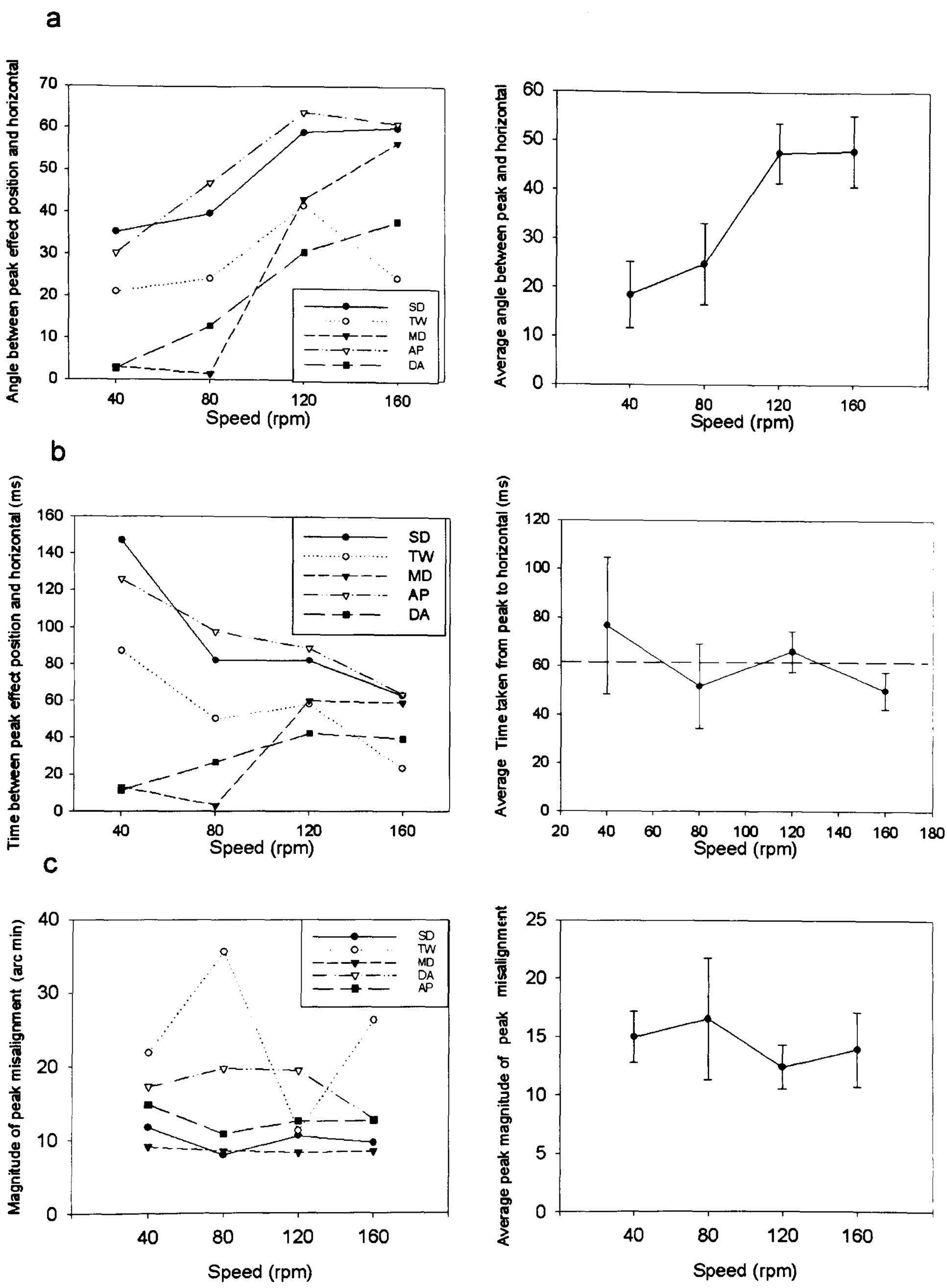


Fig. 2.5 Summary data for the fitted peaks of distributions of misalignments over all subjects. (a) The angle between peak misalignment and horizontal increases over the five subjects with speed shown along with the average angle. Error bars ±1 S.E. (b) The times taken for the

rotating bar to travel from the angle of peak misalignment to the horizontal, plotted along with the average time. Error bars ±1 S.E. There is no clear pattern over the three subjects. The average of all measurements is roughly constant around 62 ms over all speeds. (c) The magnitude of the peak perceived misalignment. There is no effect of speed overall. Error bars ±1 S.E.

2.4 - Experiment 3: Introducing background flicker

The action of motion on the target suggests the involvement of extrastriate motion selective cells with large receptive fields. In the third experiment we introduced dynamic noise into the area containing the stimulus by adding flickering noise dots in the background as a means of attenuating the influence of motion (Churan & Ilg, 2002).

2.4.1 Methods

Stimuli were all the same size and the speed of rotation of the bar was the same as in Experiment 1. The background was grey (19 cd/m^2). The experiment took place in dim ambient light with a chin-rest. The black flashed bars were presented at 1 degree separation from the central rotating bar and perceived misalignment between them about the horizontal was measured as in Experiment 1. A white fixation point was provided in the centre of the bar. The rotating bar was presented for half a rotation (0.8 s) from vertical. The flashed bars were presented once at the angle of maximal effect (60°) as found in Experiment 1. Subjective alignments and standard errors were calculated from the average of four estimates for each condition (60 trials per each alignment measurement). For the static dots condition on average 314 white (53 cd/m^2 , 5×5 arc min) dots were presented continuously during each trial (all within a circle of radius of 4 degrees containing both the bar and flashes), at different randomly assigned locations. For the temporal frequency conditions,

the dots were flickered on and off synchronously according to a square wave function at 20 Hz, 10 Hz, 5 Hz and 2.5 Hz rates and the misalignment measured for each condition. The high contrast black moving bar always occluded the background. One of the authors (SD) and four naïve subjects participated.

2.4.2 Results

We found that static noise dots had no significant effect, indicating that the shift is robust with respect to the presence of a local spatial reference, $t_4 = 1.69$, $p = 0.166$. However, although there is again some variability between subjects in the size of the effect, perceived misalignment (see Fig. 2.6) was reduced as the rate of flicker increased. We treated the data as 4 conditions $\times$ 5 subjects factorial design, and found an effect of flicker, $F_{3,60} = 19.35$, $p < 0.05$. There was also an interaction between subject and flicker rate, $F_{12,60} = 6.963$, $p < 0.05$, indicating that the decrease is multiplicative rather than a constant size over subjects. The flickering dots did not appear to mask the motion of the bar although spatial misalignment was much reduced. This suggests activating transient mechanisms interferes with the effect of motion at a distance.

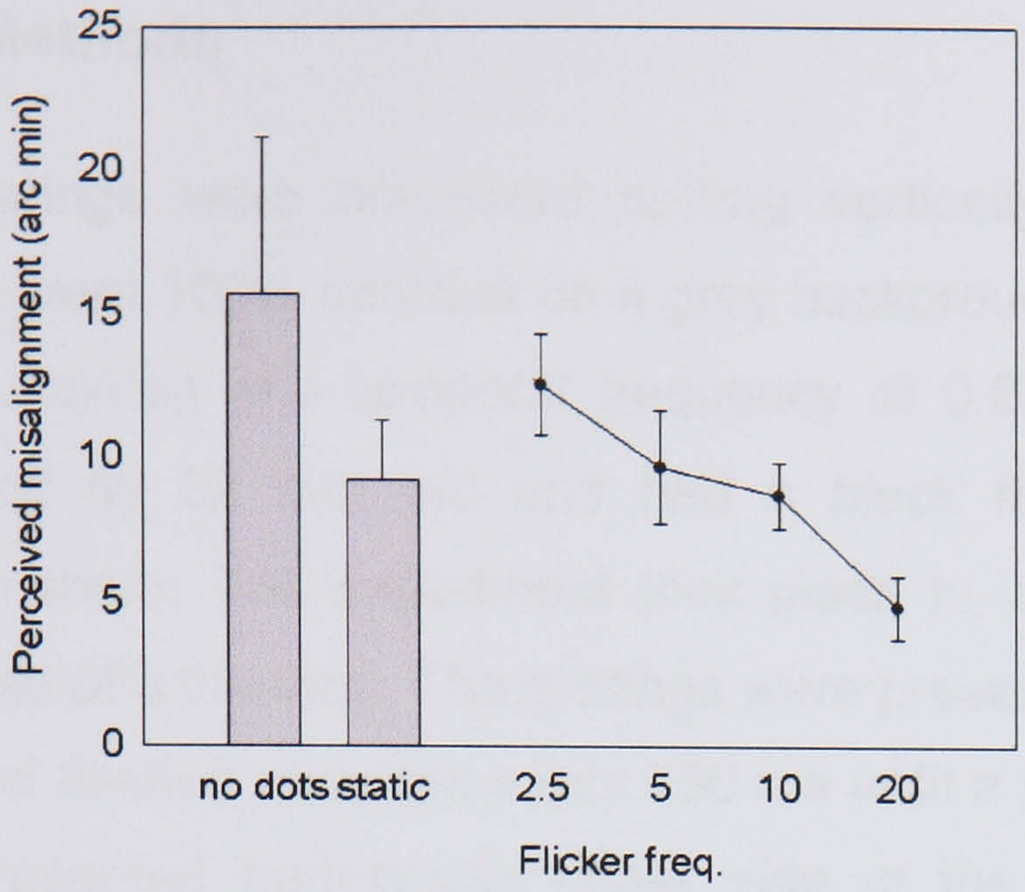


Fig. 2.6 Averaged results of Experiment 3. Error bars are +/- the mean S.E. of the subjects, to illustrate the average error for each subject, rather than error over all subjects. There is no significant difference in perceived misalignment when static dots are presented in the background. However the size of the perceived misalignment decreases significantly when the dots are flickered. The spatial shift becomes more disrupted with higher rates of flicker.

2.5 - Experiment 4: Separating the effect of eccentricity and motion distance

Whitney and Cavanagh (2000) found that motion influences position with no effect of distance between moving stimulus and flashes, suggesting a higher-order binding effect, rather than an effect of local motion. However we observed that for a given speed of rotation relative position determines the size of the effect. Previous work on motion influence on positional judgments has shown that the effect size can increase with peripheral viewing (De Valois & De Valois, 1991). In Whitney and Cavanagh's (2000) experiment increased distance from the inducer was correlated with an increase in visual eccentricity. In this experiment we separated the influence of motion over distance from retinal eccentricity. For this we used a stimulus previously utilised by Whitney and Cavanagh (2000), but manipulated the stimulus configuration so that separation from motion varied independently of the eccentricity of the flashes (Fig. 2.7).

2.5.1 Methods

Two gratings were presented drifting vertically in opposite directions. The gratings were 100% contrast on a grey background and had a spatial frequency of 2 cycles/deg and temporal frequency of 0.85 cycles/s. The gratings were separated by 32 min arc and had a black fixation point (14×14 min arc) between them. The experiment took place in low ambient light and subjects made use of a chinrest. The gratings were presented for 850 ms before the first flash and flashes occurred every 850 ms until a judgement was made. Flashes were presented horizontally either side of the gratings on an arc of equal eccentricity (5 deg 2 min arc) from the fixation point. The perceived misalignment was measured at 3 deg, 4 deg and 5 deg horizontal distances from the midline. There were 90 trials per measurement and four measurements were averaged to determine the misalignment at each distance. The author (SD) and three naïve subjects participated.

2.5.2 Results

We treated the data as a 3 conditions × 4 subjects factorial design and found a significant effect of distance from motion, $F_{2,36}=51.59$, $p<0.05$, and again found an interaction between subject and distance from motion, $F_{6,36} = 19.64$, $p<0.05$. The averaged data is shown in Figure 2.7.

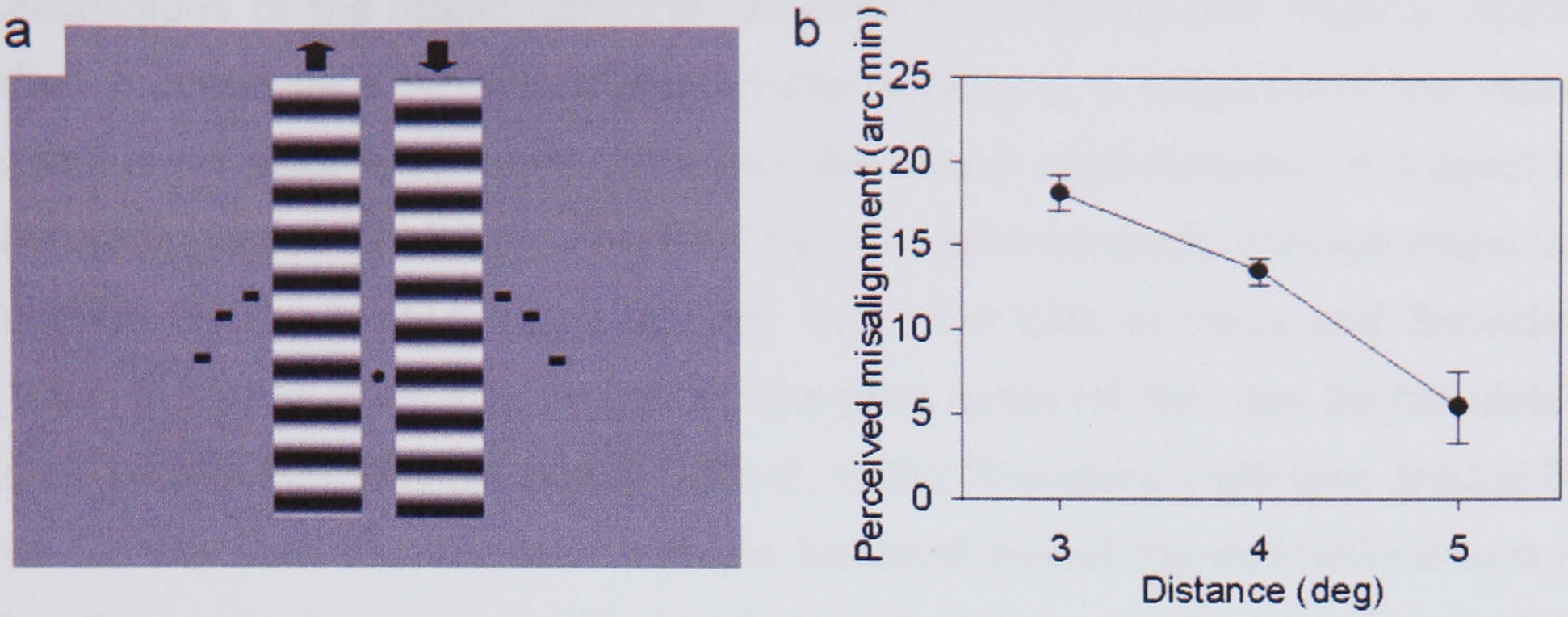


Fig. 2.7 Effect of distance from motion on perceived misalignment. (a) The configuration of the stimulus, with the three possible positions of the flashes. (b) Perceived misalignment is plotted against distance of flashes from motion. Error bars are +/- the mean S.E. of the subjects.

This data demonstrates that by controlling for the eccentricity of the flashes there is a decrease in the size of the perceived misalignment as the flashes are placed further away from motion, indicating that the extent of the influence of motion on spatial position is spatially localised and stronger the closer the flashed objects are to movement.

2.6 - Discussion of experiments

We showed that the local motion of an object can influence the perceived position of a spatially dissociated flashed static object. Using a rotating bar allowed us to examine the spatio-temporal dependence of the displacement effect. Whitney and Cavanagh (2000) showed, using their direction reversing inducing stimulus, that it is the direction of motion present around 200 ms after the presentation of the bar that determines the direction of the spatial displacement. We have shown that it is the presence of motion on the path to the flash location over a period of 60 ms after flash onset that maximizes the magnitude of the perceived spatial displacement.

Discussions of the spatio-temporal localisation of flashed and moving objects often proceed as if the visual system has access to a snapshot of the visual stimulus on each time frame. However the neural representation is subject to the spatio-temporal blurring specified by the spatio-temporal impulse response function for the human visual system. From the data of Hess and Snowden (1992) the peak latency for a realistic low-pass temporal filter can be calculated to be around 80 ms (Johnston & Clifford, 1995). Therefore it will take around 80 ms for the flash to maximally activate low-level neural representations and in this time span its position will already be affected by motion. We can explain this data if the spatial encoding is influenced by cells centred on the flash with large enough receptive fields to be activated by distant inducing objects concurrently with the neural response to the flash. We can think of a spatio-temporal window around the flash presentation, with the maximal shift occurring when the moving bar is spatially close and at a fixed temporal interval from the peak of the response to the flash. In order to arrive at the right place at the right time a faster moving bar will need to 'set off' from a more distant spatial position (Fig. 2.8). The right place would appear to be around the horizontal and the right time, 60 ms after the flash. Since motion selective cells with large receptive fields are located in extrastriate areas such as MT and MST but fine position judgments are likely to require the precision of V1, the position shift is likely to result from feedback connections from extrastriate to striate cortex (Nishida & Johnston, 1999; Whitney & Cavanagh, 2000). Motion analysis and feedback to a cell encoding spatial location will take time. Thus we need to include a small delay to account for the time it would take for feed back to influence V1 spatial codes (Fig. 2.8). Adding a delay of 20 ms for this process would provide feedback at 80 ms from the onset of the flash i.e. when we would expect the response to the flash to peak.

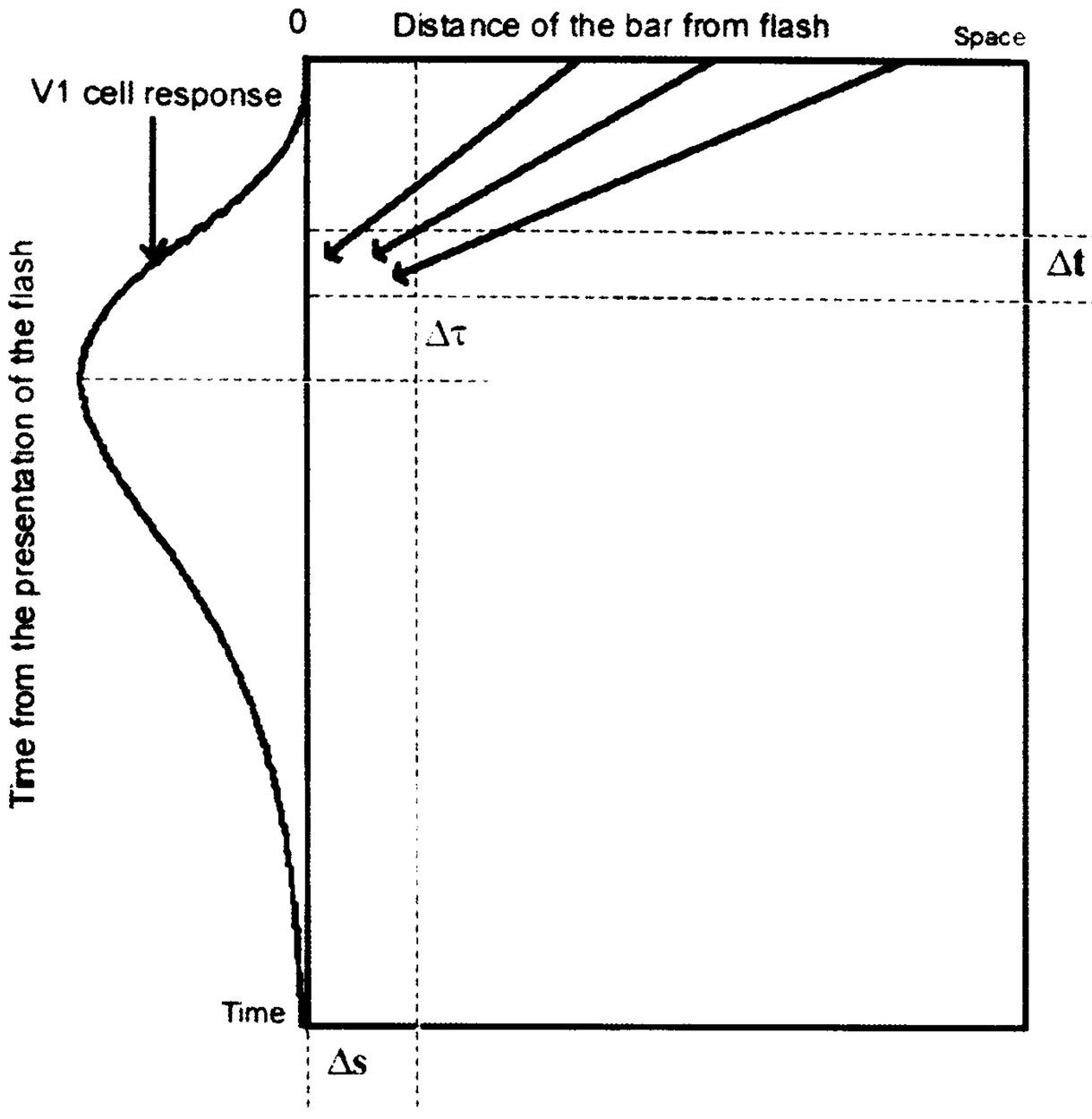


Fig. 2.8 A space-time plot depicting the traces of the bar moving at different speeds as a function of angle from the horizontal (flash location) and time from the flash. Faster moving bars need to start further from the horizontal if they are to reach the spatio-temporal window (Δs, Δt) of maximum influence 60 ms after the flash. We also need to include a motion calculation and feedback interval ($\Delta\tau$) giving a total delay of ($60+\Delta\tau$) to match the peak development of the temporal impulse response of the flash.

The stimulus used resembles the configuration used by Nijhawan (1994) to measure flash-lag, but in this perceptual alignment experiment subjects are not asked about the relative positions of the bar and the flashes. The 24 ms flash-lag we measured (separately from the spatial shift - as described in Experiment 1), is smaller than the typical temporal offset for the spatial shift. Nevertheless, it might be suggested that we have obtained an implicit measure of the flash-lag effect.

This proposal raises some interesting issues. For instance in temporal explanations of the flash-lag effect such as the differential latency model (Mateeff & Hohnsbein, 1988; Patel et al., 2000; Purushothaman et al., 1998;

Whitney et al., 2000; Whitney & Murakami, 1998), the position of the flash is established after a delay and then compared to the new position of the moving bar. However, if the flashed bar is simply delayed by 60 ms relative to the moving bar, there should be no opportunity for the moving bar to influence its position, since the moving bar would be closest to the flash when the flash first activates its neural representation. The flashed bar should initially appear in its proper retinotopic location. This does not occur as it has previously been found that for durations longer than 120 ms a flashed bar does not appear to move (Whitney & Cavanagh, 2000) but still appears spatially displaced.

The spatial extrapolation model (Khurana & Nijhawan, 1995; Nijhawan, 1994, 2002) proposes that we extrapolate the position of moving objects to correct for neural delays. One might argue that the moving bar is shifted forward by 60 ms and therefore is perceptually aligned with the flashed bar at flash onset. However we would need to extrapolate not only the position of the bar, but also the motion field, since it is the motion not the bar position that influences the flashed bars (the effect does not reverse after the bar passes the horizontal). This goes further than current extrapolation theory.

Further explanations of the flash-lag suggest that it is the side effect of a mechanism invoked to decide on a given relative spatial position at a given time for a moving object. The location of a moving object could be determined by a slow average of relative position over time (Krekelberg & Lappe, 2000, 2001), or positional sampling (Brenner et al., 2001; Krekelberg & Lappe, 2000, 2001) or by post-dictive position integration after the flash presentation (Eagleman & Sejnowski, 2000, 2002). These theories do not bear on the spatial shift effect since subjects are never asked about the relative position of the flash and the moving bar.

The influence of motion was dramatically reduced by the introduction of flicker in to the background. This is an indication that flicker can counteract the influence of motion on spatial localisation.

It has been shown that similar motion induced spatial shift effects increase with greater eccentricity (De Valois & De Valois, 1991). The decrease in the size of the shift with distance from motion described here implies a local effect of motion since we ensured that the flashes have a constant eccentricity.

The mechanisms that could underlie such a feedback mechanism have yet to be specifically proposed. The challenge for the rest of this work is to gather what we have discovered from psychophysical evidence, combine it with what is known from physiology and propose a model of spatial representation and motion processing that incorporates this information and begins to untangle the processes that are taking place.

Chapter 3- Modelling the contrast sensitivity of V1 cells

3.1 - Introduction

If one is to develop a model based on the properties of neural processes in area V1, one first of all needs to consider individual physiological cell data. The functions that describe the spatial properties of these cells must necessarily be the building blocks of any visual model.

In this thesis a model of spatial representation is proposed based on a Taylor jet representation (Koenderink & van Doorn, 1987). The components of this model are derivatives of Gaussians. This representation is proposed as a description of activity in V1. In order to strengthen this model, evidence is needed that such a derivative of Gaussians based model provides a good fit for V1 responses. Previously, examples of good fits using derivatives of Gaussians have been shown (Georgeson & Freeman, 1996; Young, 1985, 1986), however other suitable models have also been successfully used to fit physiological data (Jones & Palmer, 1987; Ringach, 2002). In this chapter I examine a past example in which several models were compared and the derivatives based approach was ruled out (Hawken & Parker, 1987). It would be hard to proceed with the proposed model at this point if we were to accept that a derivatives of Gaussian based model cannot describe single cell behaviour in V1. The aim of this chapter is to question this finding in order to further evaluate evidence for the derivatives approach and also to tackle some of the interesting questions that arise when attempting to fit models to data.

Physiological recording from single cells along the primate visual pathway has allowed us to build up detailed maps of cellular responses to light across the visual field (Hubel & Wiesel, 1962, 1968, 1974). This empirical investigation has been coupled with the development of the idea of the visual system as an information processor, designed to accentuate key features of the world around us in an efficient manner (Marr, 1982). As we gather more information on the machinery that implements these processes, we need to ask what functions we can infer from the properties of cells, so that eventually we can draw conclusions about the processing of the whole system from the individual cells. As with many branches of biology we are at a point where we find an overload of detailed biological information in need of organisation within a theoretical framework. This is when the tools of mathematics and computation can be usefully applied to find patterns and interpret results. In turn, new predictions from mathematical models open up new areas of investigation for neurobiologists.

I will now describe such an attempt at modelling physiological data using mathematical functions. This is a clear example of the challenges faced by biological modelling. We know a great deal about properties of V1 cells, including individual responses to variation in the spatial frequency of visual patterns (De Valois et al., 1982). Yet, measuring the response curve of a cell to a certain stimulus does not allow us to specify the function of the cell. We describe here past attempts to find a mathematical model to describe the spatial contrast sensitivity functions of simple cells and past conclusions about the functional role of cells in the V1 area. In particular we investigate a case where commonly applied functions have failed to successfully model data, leading to the need for new approaches. In the light of new evidence indicating that the spatial receptive fields of simple cells can be constructed from the linear combination of two even and odd-symmetric components (De Valois et al., 2000), a derivatives of Gaussians based model is introduced, which incorporates multiple channels for image filtering.

3.2 - The contrast sensitivity function

The properties of the spatial receptive field of a single cell might be determined from recording the firing rates of that cell to a bright spot or bar of light placed in its receptive field (Hubel & Wiesel, 1962). An alternative way of finding out about the cell's properties that could also lead to some deductions about its spatial receptive field is to determine the contrast sensitivity function of the cell using sine gratings that cover the receptive field (De Valois et al., 1982). In this case different sine gratings of different spatial frequencies and different levels of contrast are presented. For each spatial frequency the threshold contrast level at which the cell begins to fire is recorded. This results in a curve of contrast sensitivity (1/contrast threshold) versus spatial frequency. This method has the advantage of ease of measurement for both simple and complex cells. The shape of the receptive field is reflected in the spatial frequency tuning of a cell. Using this method, one can plot the bandwidth over which a cell responds to spatial frequency and the point at which it responds maximally.

If we can assume the properties of the cell we are measuring is linear w.r.t. to its input (i.e. doubling the contrast of the image, doubles the response of the cell), then moving from the spatial domain of the cell response to the spatial frequency domain is equivalent to taking the Fourier transform of the function of the cell's receptive field. Indeed it has been found for some cells that the Fourier transform of the function approximating the spatial receptive field is a good approximation of the shape of the contrast sensitivity function of that cell (Movshon et al., 1978). It is crucial to point out that this mathematical property only holds when a cell acts linearly on the image. As mentioned in Chapter 1 not all cells fit clearly into the 'simple' and 'complex' categories, so we have to be careful in our predictions of receptive fields even for cells that we have classified as simple. It is not true even for simple cells that they are entirely linear, as a consequence of the fact that there is no such thing as negative firing rate, therefore firing rate can increase more easily above the resting firing

rate, it can decrease much below. This can in some ways be seen as a half-wave rectification in effect introduces a non-linearity into the system (De Valois & De Valois, 1990).

There exists a psychophysical correlate to the single cell contrast sensitivity functions discussed above. One can present similar stimuli to those used in single cell measurements to human subjects and obtain a threshold measure of contrast sensitivity while varying spatial frequency (De Valois et al., 1974). Threshold is the point where the contrast is sufficient for the subject to be just able to distinguish a grating of a given spatial frequency. One of the possibilities opened up by mathematically modelling single cell data is finding the links between single cell behaviour and the overall population behaviour that leads to psychophysical observations.

3.2.1 Finding a suitable function

Mathematical functions used to model biological situations need to be constructed from biologically plausible mathematical operations. The motivation for mathematical models is often based on the linear model of cell receptive fields. In this way the Fourier transform of a suitable model of the spatial receptive field can be used to fit single cell data. However, in this case, implications of this assumption for the prediction of complex cell response need to be considered, as we are no longer dealing with a linear operation.

The value of modelling the single cell spatial contrast sensitivity function lies in the ability to then use mathematical equations to predict the response of V1 cells to different spatial frequencies and contrast, and to provide us with useful parameters of biological importance, which can provide a basis for comparison of cell behaviour, as well as allowing the prediction of simple cell spatial receptive fields.

3.3 - Past models for single cell contrast sensitivity function

The shape of the spatial receptive field along the axis orthogonal to the preferred orientation of a simple V1 cell typically takes one of the following four shapes: a central peak at the excitatory central region flanked by dips at the inhibitory regions, or vice-versa, or the odd symmetric form where we simply see an excitatory peak followed by an inhibitory dip (or vice-versa). I now introduce some past models that make use of mathematical functions that take on this range of shapes. The cell response is a function of x, the distance moved along the direction orthogonal to the preferred orientation. By assuming linearity of the cells and an estimate of Fourier phase one can use the Fourier transforms of these spatial sensitivity profiles to model their spatial contrast sensitivity.

The Gabor Model: A Gabor equation in 1D is a sine curve bounded by a Gaussian envelope and so takes the form:

$$A\sin(fx + p)e^{\frac{-x^2}{\sigma^2}} \tag{3.1}$$

with parameters

f frequency of the sine wave

p phase of the sine wave

σ space constant of the Gaussian

A scaling constant

This function has been widely used because of its close resemblance in shape to typical V1 receptive fields, its biological plausibility and because it is well defined at all points. It is biologically motivated by the fact that it is optimal in terms of compactness if one wishes to express an image in terms of space and

spatial frequency and hence is potentially significant for efficient processing (Marcelja, 1980; A. B. Watson, 1983; Hawken & Parker, 1987).

The derivative of a Gaussian model: This model is based on a differential of a Gaussian of some order, with the most commonly used variant being the second order differential, which in a circularly symmetric 2D form is the Laplacian operator, so the model can take the form $\nabla^2 G$. The Laplacian is given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{3.2}$$

In one dimension this simply becomes the second order differential of a Gaussian. In one dimension we have the following form for these functions:

$$A\frac{d^n}{dx^n}e^{\frac{-x^2}{\sigma^2}} \tag{3.3}$$

with parameters

n order of differentiation

σ space constant of the Gaussian

A scaling constant

The second order derivative takes the form:

$$A\frac{d^2}{dx^2}e^{\frac{-x^2}{\sigma^2}} = A(\frac{4x^2 - 2\sigma^2}{\sigma^4})e^{\frac{-x^2}{\sigma^2}} \tag{3.4}$$

In machine vision the Laplacian (Eqn. 3.2) function is often convolved with a 2D input image. The resulting 2D function can be used to find the points at which the second derivative of the light intensity crosses the zero axis (zero-crossings) to provide a quick methodical way of detecting sudden intensity changes, i.e. edges between light and dark areas of an image (Marr, 1982). It is possible that our visual system may be using cell receptive fields to perform a

similar task. Gaussian derivatives have been suggested as descriptions of simple cell behaviour (Young, 1986). An advantage of this model would be that the information contained in this kind of differentiated output could prove useful in further models of other forms of processing such as blur analysis (Georgeson, 1994) and motion detection (Johnston & Clifford, 1995; Johnston et al., 1992).

Differences of Gaussians (DOG): These are known to closely approximate the shape of a second order differential of a Gaussian under certain circumstances, but have also been introduced as an explanation in their own right for the shape of spatial receptive fields. A similar shape to that of the second order differential of a Gaussian can be achieved by subtracting a Gaussian from another. Models based on this can be extended to consist of the differences of a number of Gaussian terms, each specified by its own scaling factor and space constant. The simplest form, the difference of two Gaussians in 1D takes the form:

$$Ae^{\frac{-x^2}{\sigma_1^2}} - Be^{\frac{-x^2}{\sigma_2^2}} \qquad (3.5)$$

with parameters

A, B scaling constants

σ_1, σ_2 space constants of the Gaussians

It can be argued that the two Gaussians can be interpreted in terms of the organization of the components of the receptive field, where the subtracted Gaussian would represent the inhibitory inputs and the other positive part would representing the excitatory inputs (Hawken & Parker, 1987; Rose, 1979).

3.4 - Curve fitting

In order to find a curve that describes the shape of the contrast sensitivity data well, and find parameters of the given function that describe a particular cell, we

have to fit a mathematical function. We allow the function to vary in its parameters until we find a curve that passes through the data points with the minimum amount of error. Error is usually measured as the differences squared, i.e.

$$error = \sum_{i=1}^{n} \frac{(f(x_i, \mathbf{u}) - y_i)^2}{n} \tag{3.6}$$

Where (x_i , y_i) are the data points and f is the function that is fitted to the data, with **u** the vector of parameters. It is important to divide by n, the number of data points, in order to be able to compare error over data sets with different numbers of data points. The aim is to minimise the error value with the choice of **u**. Defining a best fit as the curve with parameters **u**, that gives the lowest error value (as defined above) is called the method of least squares. When f is linear in its parameters (none of them are a higher order than 1), this can be done through an extension of the method of least squares used for fitting a straight line to m terms, limited only by our ability to solve m linear equations in m unknowns (Bevington, 1992). For equations that are non-linear in their parameters we need starting values and step sizes for algorithms to converge to a least squares fit. Often convergence depends upon the choice of methods for fitting (Bevington, 1992). The following fits minimize log differences, i.e

$$error = \sum_{i=1}^{n} \frac{(\log(f(x_i, \mathbf{u})) - \log(y_i))^2}{n} \tag{3.7}$$

The minimum log error value will be reported.

3.5 - Hawken and Parker

The mathematical functions described above have all been used previously to fit V1 cell contrast sensitivity data with varying degrees of accuracy. However, we now describe some results that were not fitted well by any of the models above and the proposed extended DOG model to correct for the inaccuracies.

Hawken and Parker (1987) combine experimental and modelling work in their paper. Their aim is to measure the contrast sensitivity functions from single cell recordings of neurons in the monkey striate cortex to determine which model fits these functions best and hence determine the best model for reproducing and predicting the behaviour of visual neurons. They introduce two variations on the DOG model.

The first is the DOG-S model where instead of the whole receptive field being described by the difference of two Gaussians, each subunit is described by a separate Gaussian. Two parameters describe the central peak and two Gaussians are then subtracted to form the two flanking regions. The flanking regions are constrained to be symmetric and so have the same two parameters and an additional parameter, S, the separation from the central Gaussian.

The DOG-S is a stripped down version of their proposed d-DOG-S model, which is the difference of difference of Gaussians with a separation parameter, where each component of the receptive field is described by a difference-of-Gaussians. The flanking regions are constrained to have the same space constants and amplitudes, but the symmetry of the receptive field can be described with the symmetry parameter, g. Hence, it takes the following form:

$$k_{c_1} e^{-\left(\frac{x}{x_{c_1}}\right)^2} - k_{s_1} e^{-\left(\frac{x}{x_{s_1}}\right)^2} - g(k_{c_2} e^{-\left(\frac{x+S}{x_{c_2}}\right)^2} - k_{s_2} e^{-\left(\frac{x+S}{x_{s_2}}\right)^2}) - (1-g)(k_{c_2} e^{-\left(\frac{x-S}{x_{c_2}}\right)^2} - k_{s_2} e^{-\left(\frac{x-S}{x_{s_2}}\right)^2}) \quad (3.8)$$

The authors measure the responses of both simple and complex neurons and classify them according to whether they display a linear response or not. The contrast sensitivity values at different spatial frequencies are determined by psychometric methods, from cell response recordings of cells in the foveal region of the primate striate cortex using micro-electrodes. Each data point is the mean of 12 estimates from a staircase procedure, with error bars shown one SD from the mean.

The non-linear fits employ an algorithm that minimizes Σ (log model – log data)^2 , the log least square measure. The fitting algorithm is not specified and convergence criteria are not specified.

For two of the cells that they consider to be simple cells they use the parameters from the fitted spatial contrast sensitivity function to derive - using its Fourier transform – the theoretical spatial receptive field of the cell. For these two cells they show the best-fit results for all the models (Fig. 6-10 from (Hawken & Parker, 1987)) and then show further examples (Fig. 13 from (Hawken & Parker, 1987)) of the variations on shape of the contrast sensitivity functions and the versatility of their proposed d-DOG-S model.

3.5.1 Results found

This paper provides a wealth of information and data that allows the comparison of models via curve fitting. The contrast sensitivity data for each cell are shown on logarithmic axes, with standard deviation bars, as each data point is actually a mean of a sample of measurements. There are 7-14 data points on each graph. Mostly we see (with some exceptions) the upside-down U-shape associated with the band pass property of single cell spatial contrast sensitivity functions. However within these we also see a great deal of variability in shape and bandwidth and also a few low pass shapes occurring (showing response to low spatial frequencies). An interesting feature that reoccurs in many of the graphs is a small “shoulder” in the data towards the high frequency end of the upside down U-shape, this is not noted by the authors as significant, however it does appear in the shape of curves that they find fit the data best.

The failure of the Gabor and 2nd differential models to capture the shape of the data is clearly illustrated. In the best fits that they show, these models both peak at a different height and frequency to the data and are less symmetric, and so fail to fit the low frequency limb of the data. These models also seem too smooth to reflect the irregular shape of the data. The calculated error per

data point for each fit is found to be significantly higher for these two types of curves than the d-DOG-S model.

The d-DOG-S model appears to perform the best out of all the models, clearly appearing to reflect the shapes of the data curves the most accurately. The authors draw the conclusion that because of this, their model is in fact the best for representing the spatial contrast sensitivity function of single cell data. This implies that each sub-region of the receptive field is represented by a difference-of-Gaussians function and that 9 parameters are necessary to define the contrast sensitivity function for each curve accurately. Although the DOG-S and DOG models also fit the curves well, the paper claims these are simply good approximations of d-DOG-S, the model that they put forward as underpinning early visual processing.

3.6 - Problems with non-linear curve fitting

The paper does not go in to detail about the algorithm that they use for their curve fitting, but this is where the first problem arises. Non-linear curve fitting is a complicated process for which there are no sure-fire methods of finding a "best-fit". Their assumption that they have found the best fit for each model is subject to a number of caveats.

1. Fitting criteria: Although the Gabor and second differential model don't fit as well on inspection by eye, one of the features the authors emphasise is the mismatch in the height of peak spatial frequency, and this is misleading. The height and position of the peak depends to some extent on the criteria used for fitting. By minimising the squares of log differences the peaks tend to be lower. If we use conventional least squares as our measure we get slightly different results than those presented in the paper (Fig. 3.1).

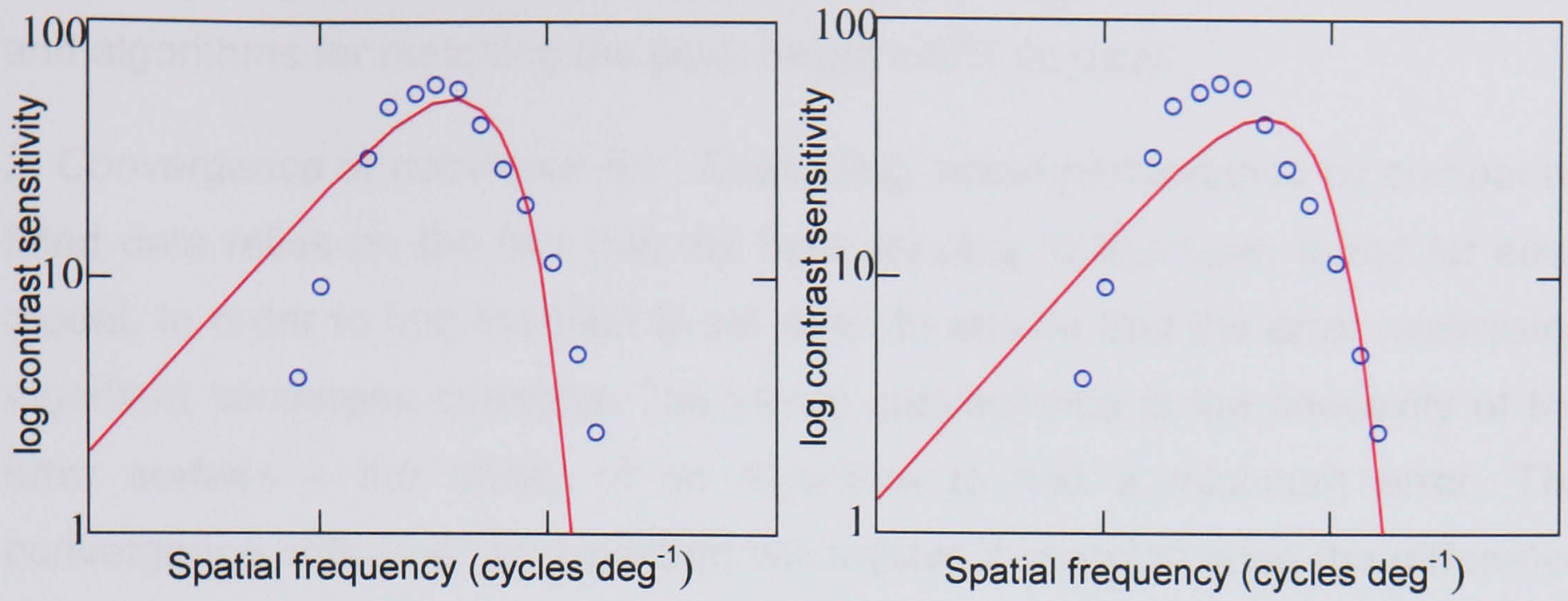


Fig. 3.1 Data points taken from the first figure in Figure 6 (Hawken & Parker, 1987). On the left is shown the curve fitted using the genfit routine from Mathcad with the Gabor model, minimising difference of squares rather than difference of log squares. With parameters $A=114.9$, $f=3.02$, $p=0$, $\sigma=0.302$ in the Fourier transform of Eqn. 3.1. On the right is the Gabor fit, minimising log squares, the same fit as found in (Hawken & Parker, 1987).

The peak we find in this case is much closer to the peak of the data. Also, we can express these data on linear axes:

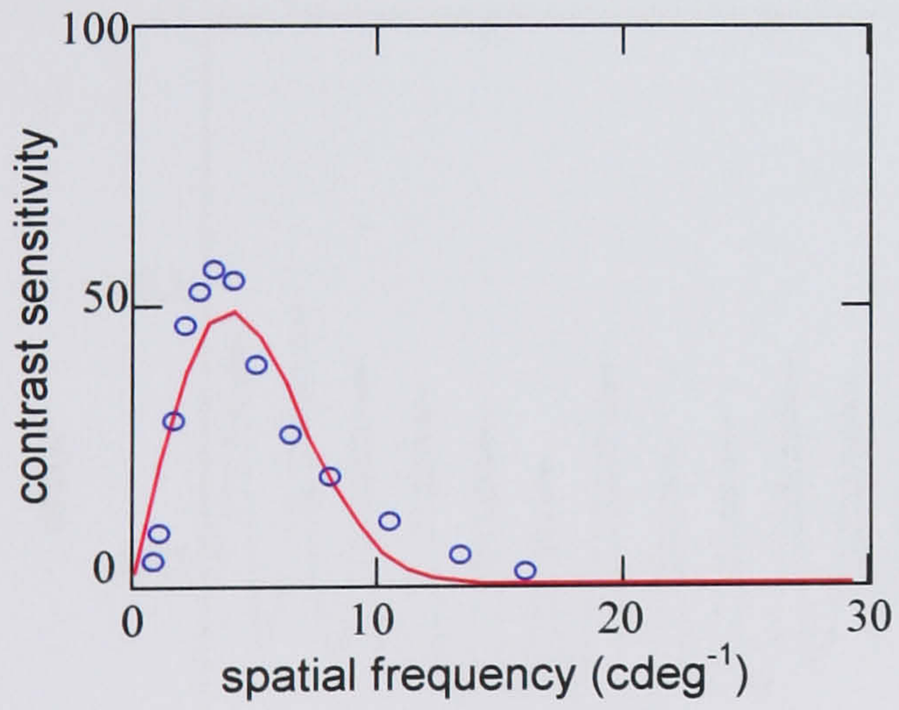


Fig. 3.2 As in Fig. 3.1, but drawn on linear scale axes

On this graph (Fig. 3.2) we see that it is the high frequency end that appears to fail to fit the data, the opposite to the impression we get when the data are plotted on the log axes. If one were indeed interested in matching peak

frequency height, then the least squares method might not be the most suitable and algorithms for matching the peak height could be used.

2. Convergence of non-linear fits: Evaluating model performance by comparing fitted data relies on the fact that the best possible fit has been found for each model. In order to find the best fit we need to ensure that the error minimising algorithm converges correctly. The key to convergence is the convexity of the error surface – the ability of an algorithm to find a minimum error. The convergence criteria of an algorithm will involve it stopping once the difference in error between each step drops below a certain number. This means we do not want the error surface to contain wide basins of equal error below this cut-off point, as the algorithm could stop at any of a wide range of points in parameter space. Another danger can arise in the case of uneven error surfaces where the algorithm might converge to some local minimum rather than the overall minimum. To illustrate the latter case we show an example of what happens when we keep all parameters fixed in the d-DOG-S model and vary only the separation parameter, S:

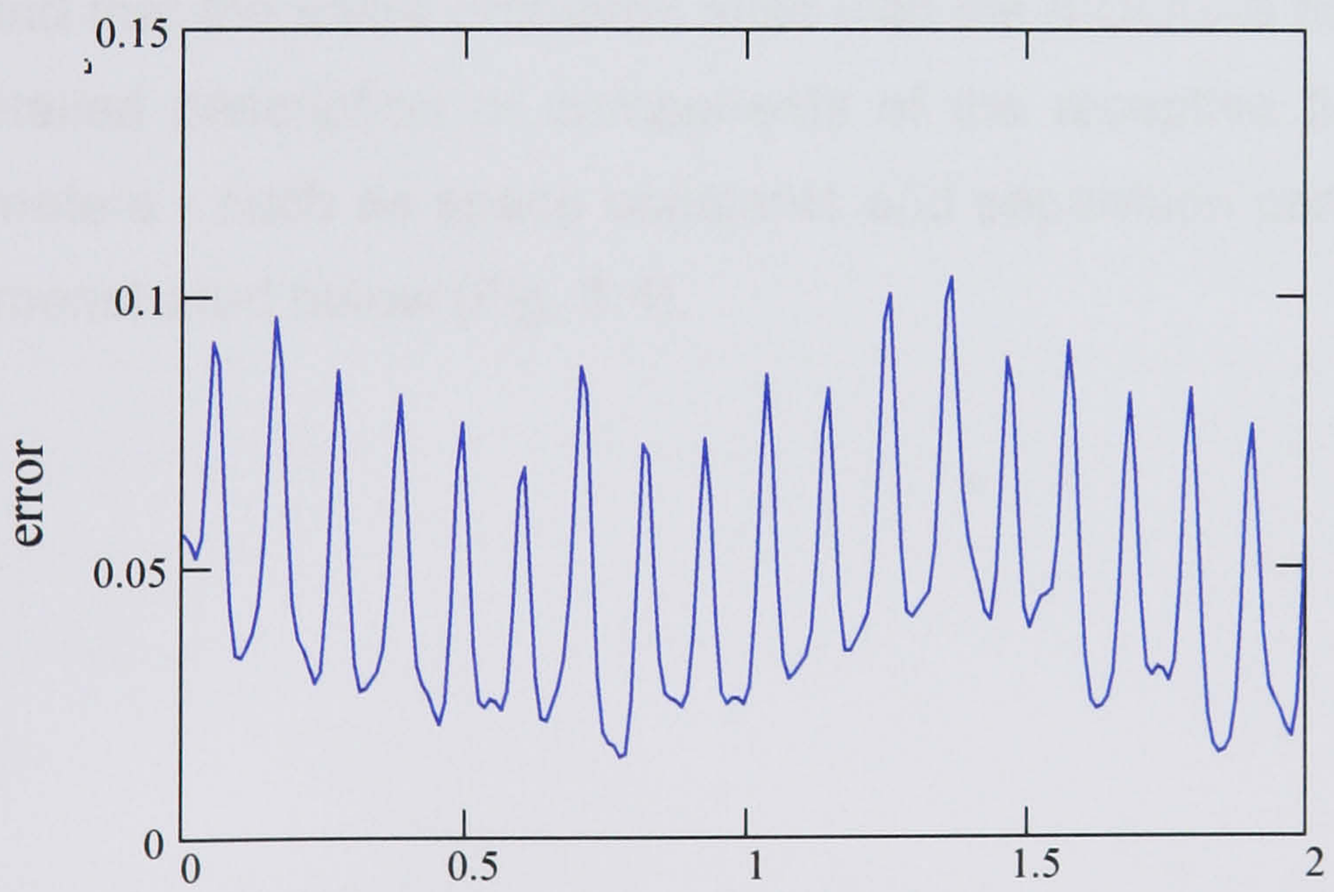


Fig. 3.3 The second data set from Fig. 6 in Hawken and Parker (1987), the error values of the data from the d-DOG-S function as the S parameter is varied.

What we see here is an example of a very uneven error surface. We can see that although in this case there is an overall minimum, the algorithm may easily find a different local minimum.

If the former case does not hold, and we have a wide range of parameters that give the similarly low error value (within the tolerance range), this means that we can end up on any number of best fits over a wide range of parameters. This means that although we may find a curve that fits the data well, its parameters are not meaningful. It is sometimes useful in this case to identify parameters that may be "trading off" against each other i.e. changing one in proportion to another does not affect the goodness of fit. Yang et al (1995) found this in their model of contrast sensitivity. They state that this behaviour would make it difficult to use their parameters in a biologically meaningful way. This behaviour could sometimes imply that it is the ratio of the parameters that is important and this would allow a pair to be replaced by a single parameter value. The points above highlight some of the problems that can arise from fitting functions that are non-linear in their parameters.

We find that the same problems arise with the d-DOG-S fits, making the claims of detailed description of components of the receptive field according to the parameters - such as space constants and separation constants - invalid. This is demonstrated below (Fig. 3.4).

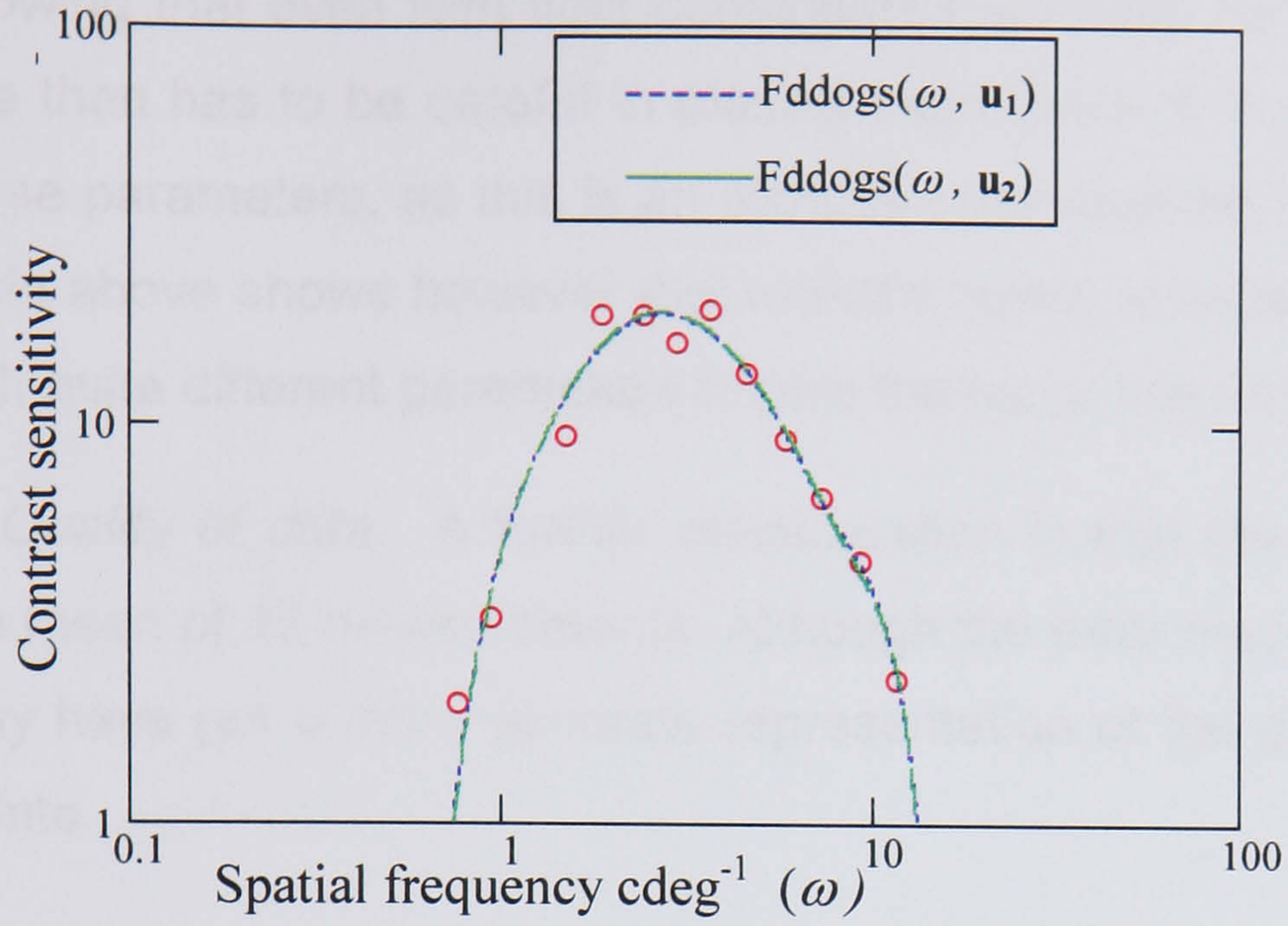


Fig. 3.4 Two different fits of the d-DOG-S model to the second set of data in Fig 6 of (Hawken & Parker, 1987) found by the genfit routine in MathCad 2000 (see beginning of section 3.9)

The parameters and error per data point for these two fits:

	kc_1	xc_1	ks_1	xs_1	kc_2	xc_2	ks_2	xs_2	S	*error*
$\mathbf{u_1}$	281.575	0.048	76.598	0.207	9.407	0.132	77.023	0.012	0.091	0.03
$\mathbf{u_2}$	283.111	0.045	71.007	0.214	33.39	0.076	66.232	0.034	0.086	0.03

We see that some of the parameters are quite different and yet the error is the same. For example, the two weighting parameters kc_2 and ks_2 give a ratio of 8.19 in the first example and 0.50 in the second, a different proportion of combination that gives the same result. The Mathcad routine can converge on different results each time, depending on starting parameters. This is a consequence of the error surface formed by the d-DOG-S model as we will discuss in more detail.

3. Constraining parameters: In the paper, to eliminate a wide range of fits with nonsensical parameters the authors constrain them, by keeping the space constants within some range based on ganglion cell measurements. The maximum sensitivity of the individual Gaussians was constrained to be within 1.5 times the maximum sensitivity of the data. This is an acceptable method of

showing that even with tight constraints the model fits the data well. However, one then has to be careful in placing importance in the biological plausibility of these parameters, as this is an obvious consequence from the constraints. The table above shows however that realistic space constants can still be combined with quite different parameters to give the same low error.

4. Quality of data: A further consideration is that the data points are actually the mean of 12 measurements. Although the error bars are relatively narrow we may have got a more accurate representation of the data by using all the data points.

3.7 - Problems with d-DOG-S model

3.7.1 Non-unique parameters

As we saw above the d-DOG-S model is not immune to the problems that befall non-linear curve fitting when trying to find a unique set of best fitting parameters. In order to examine this more closely, let us take a look at the two different d-DOG-S curves we compared that had different parameters, but gave the same error results and shapes of curves. If we vary the two parameters that are most different from each other we can build up a 3D error surface (Fig. 3.5).

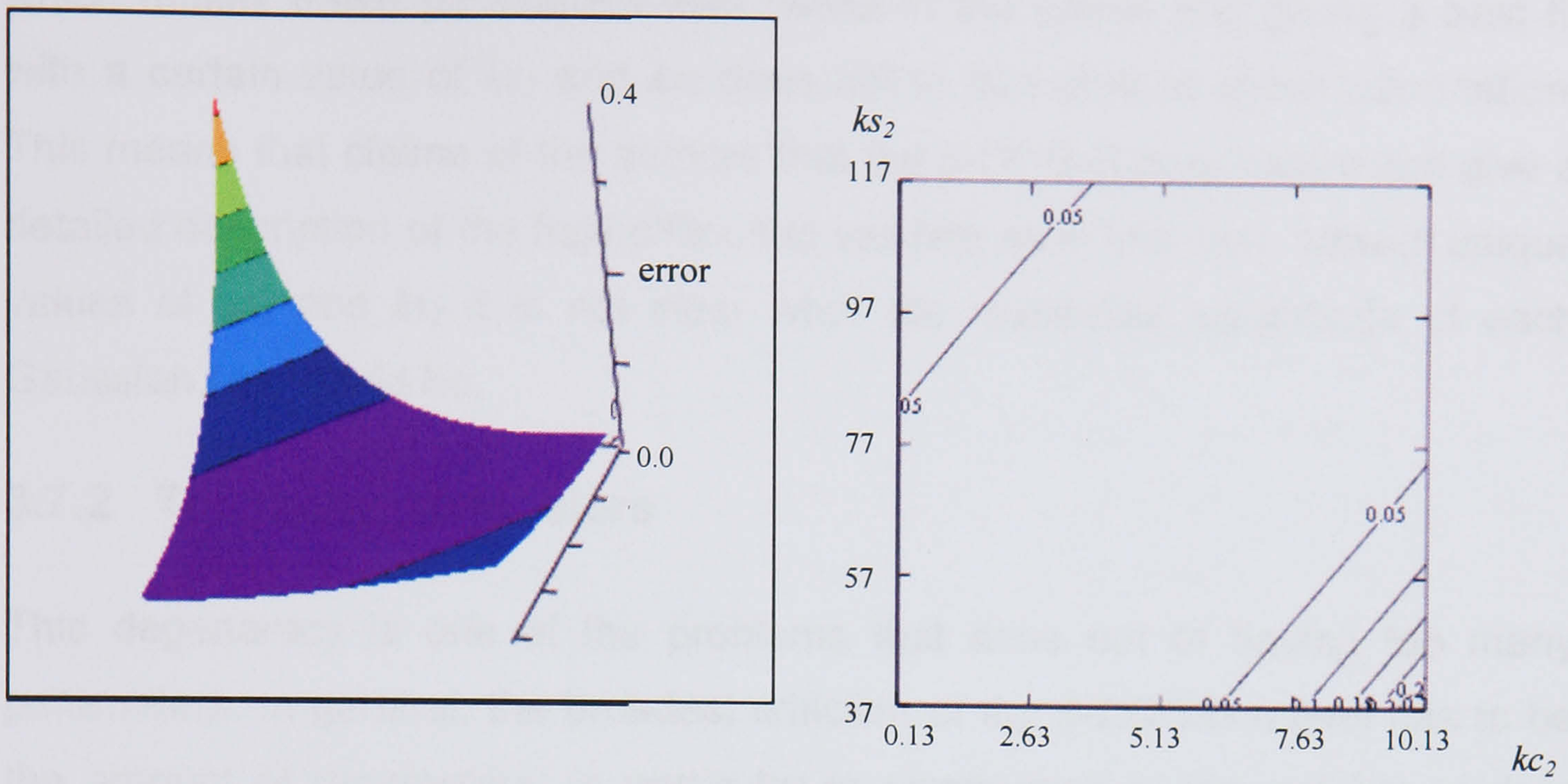


Fig. 3.5 Shown plotted along the x-axis are values of kc_2 and along the y-axis values of ks_2 with the other parameters fixed. The surface height at (x,y) shows the error per data point for the corresponding value of $kc_2=x$, $ks_2=y$. These are the error values found for the 2nd data set of Fig 6-10 (Hawken & Parker, 1987), we found good fits for kc_2, ks_2 in this region as indicated by the low error values plotted (lowest error around 0.017). The corresponding contour plot of error values is also shown .

As we can see there is a wide, flat area where the error values are all very similar even though the parameters range over a wide range. The data is plotted around the area of minimum error. The two parameters are extended in an area around the minimum until the matrix of error values contains a value of 0.4 or above (for later comparison). The horizontal lines represent level sets of equal error values. It is important to bear in mind that the shape of this error surface can be misleading, because its convexity will depend upon the scale chosen. However, we have allowed the values to vary over a wide range on the scale of these parameters and the majority of combinations give very low error values. Even keeping all other parameters fixed, kc_2 and ks_2 can range from 6-9 and 71-105 respectively with the error remaining at the lowest value of 0.019. This range of best fits lies along a line as shows in the graph in Fig. 3.5, along the line of $ks_2 \cong 11\ kc_2$. In this case the algorithmic search for a best fit could stop anywhere in this flat valley. It is this shape that leads to the fact that many

values of kc_2 and ks_2 can be chosen that would produce equally low errors, which makes these parameters less useful in the sense that giving a best fit with a certain value of kc_2 and ks_2 does not in fact give us much information. This means that claims of the authors that the d-DOG-S parameters can give a detailed description of the field difficult to validate as in this case without unique values of kc_2 and ks_2 it is not clear what the respective weightings of each Gaussian pair would be.

3.7.2 Too many parameters

This degeneracy is one of the problems that arise out of having too many parameters. In general, the broadest criticism of the d-DOG-S model has to be the amount of parameters, in particular in comparison to the number of data points. To use nine parameters to describe a data set of only thirteen points (in some cases down to even seven points) is meaningless. In general the more free parameters in a model, the more likely we are to find good fits, however, including too many parameters can lead to numerical ill-conditioning and to excessively complex models (Stark, 1997). Over-fitting often appears as a result of selecting a too complex model for the data. Given ten data points from an experiment, a 9^{th}-degree polynomial could be fitted through them exactly (Bevington, 1992). By over fitting we are loosing sight of biological meaningfulness and the information that can be extracted by applying the model. It is true that in the actual system there are probably many more factors that contribute to the behaviour of a neuron than we account for in our models, but in order to gain insight in to a system it is not satisfactory to simply replicate its complexity with an equivalently complex model.

3.7.3 Motivation of the model

Another problem is the logic that leads to using the Fourier transform of the d-DOG-S equation to fit the contrast sensitivity data. In order to move in this way from the spatial to the spatial frequency domain we have to assume that the cell is linear. However, in the paper this function is also used to fit cells that

have non-linear properties – ones we are more likely to label complex cells. Although the function still fits well, it loses its meaning back in the spatial domain. We cannot use it to infer the spatial receptive field of the cell.

Another point is the validity of using a difference of Gaussian to describe each component of the receptive field. Whilst the simple difference of Gaussian represents an inhibitory and excitatory output, supported by physiological evidence, there is less clear evidence for each subunit of the receptive field to be a result of excitatory and inhibitory inputs in the form of DOG functions. Also, both these models ultimately limit the number of lobes that can make up a receptive field.

All in all, this leads to the questioning of the biological foundation of the d-DOG-S model, the main points being in summary:

- too many parameters
- non-unique parameters
- motivation based on linear models of early visual processing neurons

These points suggest that the argument that each on/off part of the receptive field is represented by differences of Gaussians, cannot simply be shown to be correct by successfully fitting the Fourier transforms of these to contrast sensitivity functions.

3.8 - Alternative models

As stated before, although doubts have been raised about the validity of the d-DOG-S model it is clear that the Gabor and 2nd differential of Gaussian models are not suitable alternatives, as they fail to reflect the data provided in Hawken and Parker (1987) accurately. The idea for the following alternative models is based on the notion of derivative operators as spatial filters. By using various derivatives of Gaussians to implement such filters we are blurring the image

and differentiating it in one step. Evidence for multiple spatial frequency-tuned filters is found from physiological studies as well as psychophysical (Bruce et al., 1996). Psychophysical channels however, reflect both simple and complex cell behaviour, so these models are not based on the shape of the simple cell receptive field, although derivatives of Gaussians do reproduce the shape of these effectively. These channels could then be combined to produce orientation and direction selectivity. For instance in (De Valois et al., 2000) it is suggested that spatial receptive fields of directional V1 cells are a linear combination of two components that have two different spatial characteristics.

Different orders of differentials of Gaussians: First of all we simply try fitting differentials of Gaussians of different orders than 2 to the data, as multiple spatial channels could imply the presence of derivatives of higher orders (see Fig. 3.6). We would expect some simple cells to be fit by higher order differentials of Gaussians if these cells are the first step of combining different spatial derivative channels. Differentiating a Gaussian is equivalent to multiplying it by $2\pi i\omega$ in the Fourier domain, where $i = \sqrt{-1}$ and ω = spatial frequency. We take the magnitude of this complex valued function for fitting the cell data. So it holds that

$$F(A\frac{d^n}{dx^n}e^{-\frac{x^2}{\sigma^2}}) = B\omega^n F(e^{-\frac{x^2}{\sigma^2}}) = B\omega^n e^{-\pi^2\sigma^2\omega^2} \tag{3.9}$$

Where F implies taking a Fourier transform and A and B are arbitrary constants.

Fitting simple cells with these functions would imply the shape of spatial domain receptive fields. We find however, that the order of a differential does not change the basic shape of the curve in the Fourier domain in a way that resembles the data more accurately.

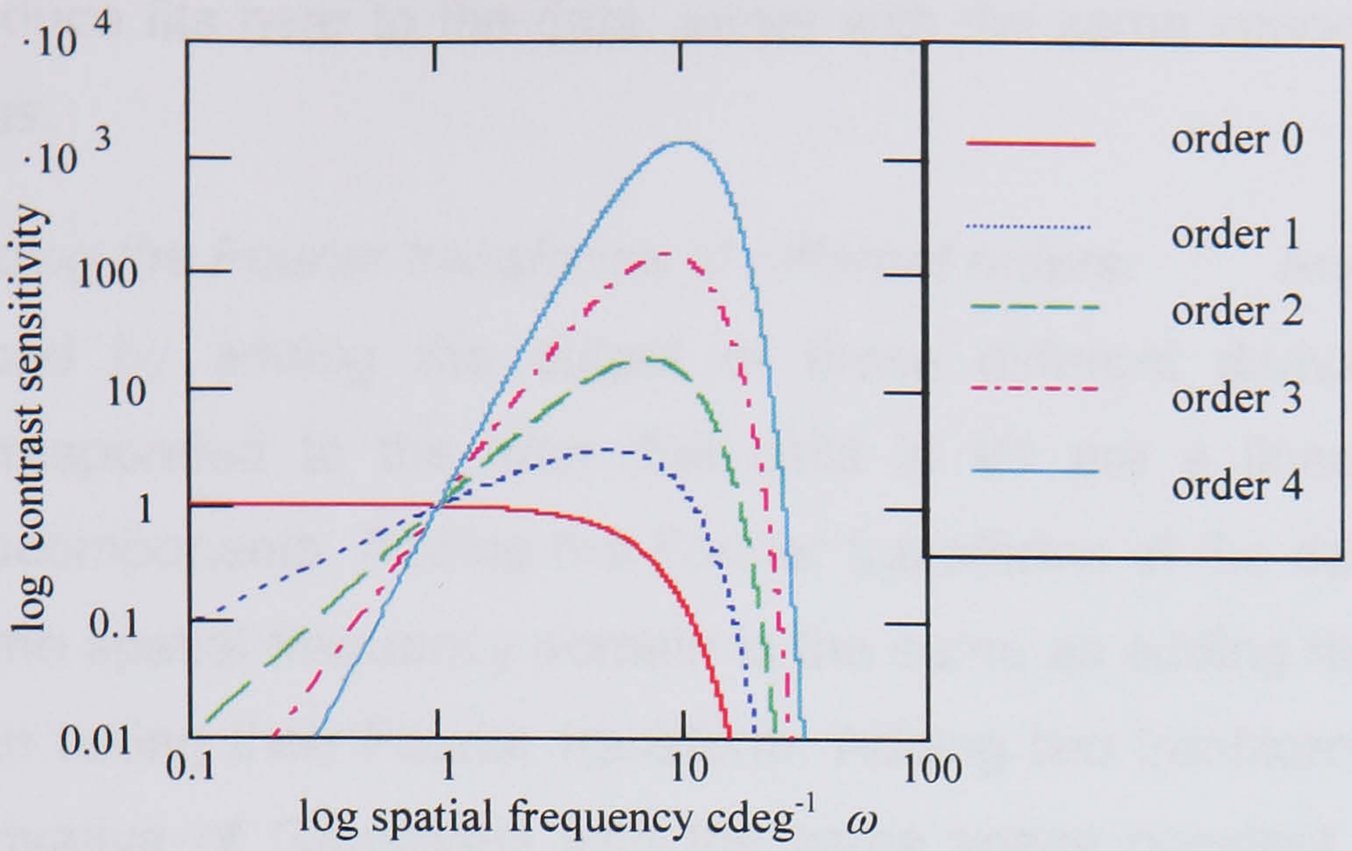


Fig. 3.6 The Fourier transforms of the differential of Gaussian ($G(x) = \frac{e^{\frac{-x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$) function of varying orders of differential, space constant $\sigma = 0.3$, drawn on log axes

All these curves have similar smooth shapes to the second differential, and fail to fit the data accurately. (Apart from a simple Gaussian transform for Fit 11, which will be discussed later).

Multiplying the Fourier transforms of different orders: We need to combine these different spatial filters in some way in order to arrive at properties such as motion direction selectivity and the cells we are examining may already be the result of such combined outputs. Two cells could combine the product of their responses to enhance large responses to spatial frequency and minimise small responses. Multiplying two differentials of a Gaussian and transforming their product, is equivalent to convolving their transforms. The Fourier transform of an *n*th order derivative of a Gaussian is: $\sqrt{\pi}\omega^n e^{-\pi^2\omega^2\sigma^2}$. Therefore, if we multiply two Gaussians with different space constants in the spatial domain we end up fitting the following function to the data in the frequency domain:

$$A(\sqrt{\pi}\omega^n e^{-\pi^2\omega^2\sigma_1^2} * \sqrt{\pi}\omega^m e^{-\pi^2\omega^2\sigma_2^2}) \qquad (3.10)$$

This is a Gaussian multiplied by a polynomial in ω (spatial frequency). Various combinations of powers of derivatives were tried in this function, but it could not

produce fits near to the data, either with the same space constants or different ones.

Adding the Fourier transforms of different orders: Another possibility was raised by adding the output of these different derivative channels, which corresponded to the idea that cells in V1 are a linear combination of two subcomponents. Adding the Fourier transforms of the derivatives of Gaussians in the spatial frequency domain is the same as adding these functions first and then taking their Fourier transform. Adding two transforms of different order of derivative of Gaussians with the same space constant leads to the function: $(A\sqrt{\pi}\omega^n + B\sqrt{\pi}\omega^m)e^{-\pi^2\omega^2\sigma^2}$, where A, B are the scaling constants of each Gaussian, m, n are the orders of the differentials, σ is the space constant and ω is spatial frequency. This multiplication of a Gaussian with a polynomial in ω cannot produce the kind of irregular shapes seen in the data.

Different space constants were then tried first by adding the two functions with same weightings, but no satisfactory fits were found. This led to the addition of two spatial derivative filters with different space constants, and with different weightings.

3.9 - Fitting the sums of Fourier transforms of derivatives of Gaussians

The motivation for this model comes from physiological and psychophysical evidence for the existence of multiple spatial channels and the idea that these perform a blurring and differentiation process that can be modelled with derivatives of Gaussians (Johnston et al., 1992; Marr, 1982). This is further confirmed by the findings that cells in V1 appear to have two main spatial components with different properties. If a cell linearly summed the input of two cells, its spatial contrast sensitivity function would be the sum of the two inputs. These two inputs may differ in order of differentiation, space constant and how

much each input contributes to the behaviour of the cell we are attempting to model. The resulting model of a cell's contrast sensitivity function takes the form:

$$sumdiff(\omega, m, n, \sigma_1, \sigma_2, a_1, a_2) = a_1 F(\frac{d^m}{dx^m} e^{-\frac{x^2}{\sigma_1^2}}) + a_2 F(\frac{d^n}{dx^n} e^{-\frac{x^2}{\sigma_2^2}}) \qquad (3.11)$$

Where *F* stands for the Fourier transform of a function.

The analytical Fourier transforms take the form:

$$sumdiff(\omega, m, n, \sigma_1, \sigma_2, a_1, a_2) = a_1 \sqrt{\pi}\sigma_1 \omega^m e^{-\pi^2\omega^2\sigma_1^2} + a_2 \sqrt{\pi}\sigma_2 \omega^n e^{-\pi^2\omega^2\sigma_2^2} \qquad (3.12)$$

(using Bracewell's (2000) definition of the Fourier transform as used by Hawken and Parker).

The 6 parameters in Eqn. 3.12 are:

σ_1, σ_2 are the two space constants of the Gaussians whose derivatives are being added

m, n are the orders of differentiation for each respective Gaussian

a_1, a_2 are the two respective weighting scalars, non-negative as the filters are added instead of subtracted.

As we can see, the number of parameters has been reduced by 3 from the 9 parameter d-DOG-S model. It may in fact be possible to reduce the number of these parameters further by means that will be described later. First of all we can note that order of addition doesn't matter, so for each best fit swapping all the parameters round between the two differentials of Gaussians (i.e., a_1 with a_2 etc.) will provide the same curve, so there is one set of good fits we can discount. Also, although in the final equation allowing n, m to take on non-integer values gives us real curves in the Fourier domain, in fact, as we are

starting from the premise that they represent orders of differentiation, we have to keep them integer. It also may be more realistic to choose these powers adjacent to each other as in this way we would be combining odd and even functions, which are the characteristics of the two types of spatial receptive fields combined in (De Valois et al., 2000). Combining curves in quadrature is an effective way of highlighting change in brightness (De Valois et al., 2000).

Method: The data was taken from the Hawken and Parker paper by scanning the graphs and taking readings of them using the Mathcad 2000 image viewer. The genfit function within the Mathcad 2000 package was used to fit the data. The routine takes a function and its partial derivatives with respect to the parameters as input, along with an initial guess for the parameters. It then returns the parameters that give the best least squares fit. The log of the function was used as input to minimise the log differences squared as in Hawken and Parker. Limiting m and n to integers at the start of fitting did not produce good convergence using genfit [www.mathcad.com]. Instead the function was fitted without limiting m and n to integers and then the two nearest integers were chosen to the best fit found. m and n were then fixed as these integers and the fitting procedure was repeated, to generate the results.

Results: (See Fig. 3.9-3.19 for graphs and further specific comments). It was found that the function in Eqn. 3.12 fitted the data as well as the d-DOG-S model, in some cases producing poorer fits and in some cases producing better ones, but reproducing in similar way the variety of shapes found in the data, also capturing the irregularity of the shapes including the “shoulder” of some of the data sets we mentioned in describing the Hawken and Parker (1987) data. All of the more typically shaped contrast sensitivity curves were fitted best by the sums of Fourier transforms of 2^{nd} and 3^{rd} order derivatives of Gaussians. All but two of the curves were fitted best with derivatives of adjacent orders. Data that wasn’t fitted by the sum of second order and third order derivatives, was

fitted by the sum of the transform of a simple Gaussian and the transform of a higher derivative. One of the cells could be fitted best by an 8^{th} and 9^{th} order derivative, but lower adjacent order derivatives also fitted it quite well. One would have to be suspicious of such a high order as this results in a high order polynomial in the function. The curve in Fig. 3.17 fitted equally as well as all the others, whereas in the original paper the DOG model worked better then the d-DOG-S. Where Hawken and Parker (1987) used a simple Guassian for the data in Fig. 3.18 the model presented here had no problem in converging to the same shape. It was found that in most cases the fitting function converged to a unique set of best-fitting parameters. The space constants were sensitive to change up to 3 decimal places, whereas the weighting constants were only sensitive up to one decimal place. These are far more sensitive parameters then those in the d-DOG-S as was shown above, with no need to constrain them to achieve unambiguous results. This is illustrated below.

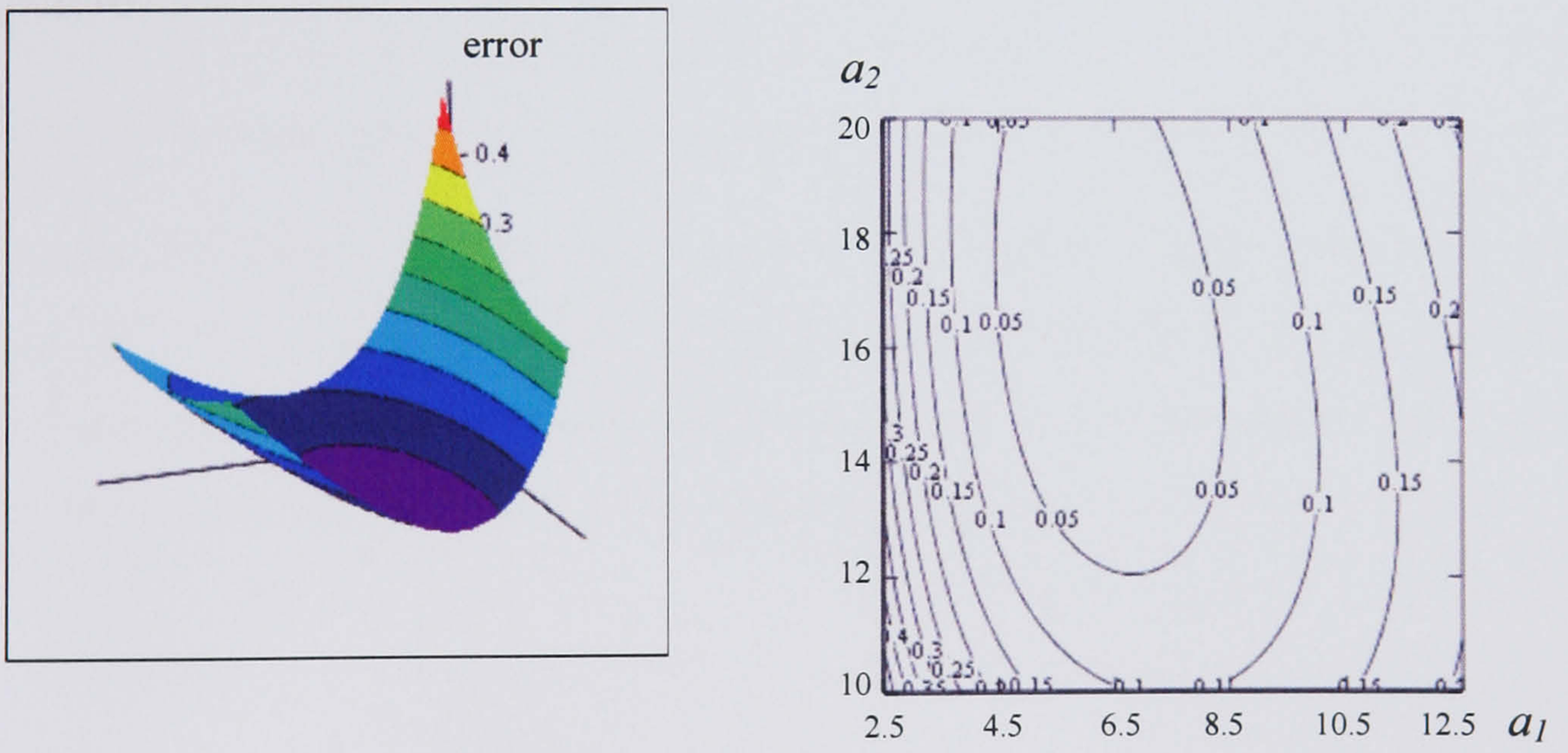


Fig. 3.7 The error plot for the sum of differentials of Gaussians model as the parameters a_1 and a_2 are varied and the corresponding error per data point is plotted on the surface for data set 2 Fig. 6-10 (Hawken & Parker, 1987). The error values are also plotted on a contour plot.

In the figure above we have along the x-axis a_1 and along the y-axis a_2, with the other parameters kept constant and the error illustrated by the surface height,

for one of the cell's data. This makes for a good comparison with the example we used for the d-DOG-S model, as these values perform a similar weighting role as ks_1 and ks_2, and also have the same order of magnitude. We show the same error surface as before, i.e. the smallest area containing both the minimum error value and the first values that give an error of over 0.4. The same level set contours of equal error are shown as in Fig. 3.5. We show the parameters over a much narrower width of range. If the ks_1, ks_2 example for the d-DOG-S model, were drawn over a narrower range it would appear even flatter. What we see is a clear well-defined minimum, with the error surface increasingly more sharply away from it then for the d-DOG-S example. If we compare the range for the minimum values of error 0.016, it varies over a_1= 6-6.2, a_2=15.9-16.7, less variation than for the d-DOG-S weighting parameters above.

For the space constants compared against each other up to similar error limits, again within a narrow range on the space constant scale, we get a similar picture:

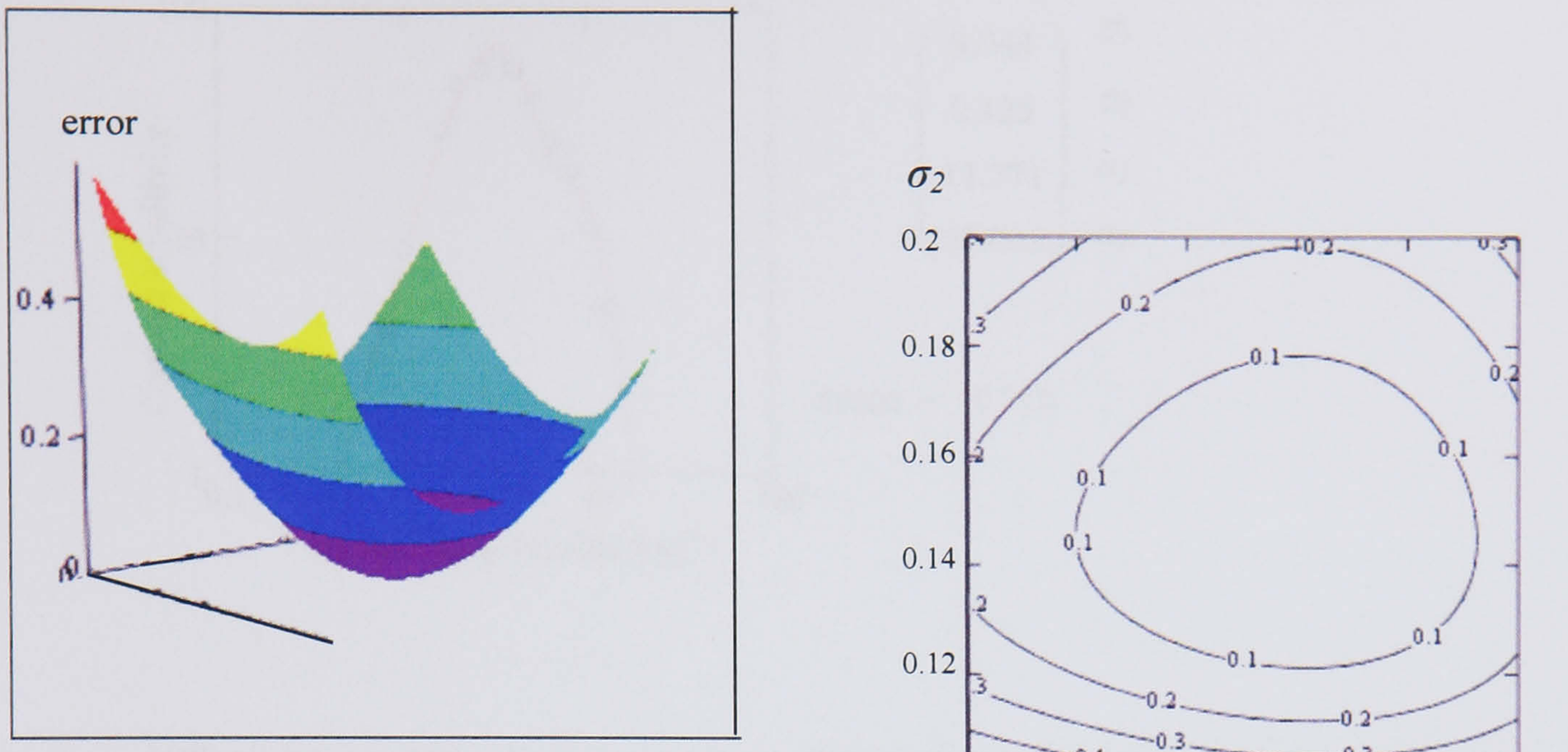
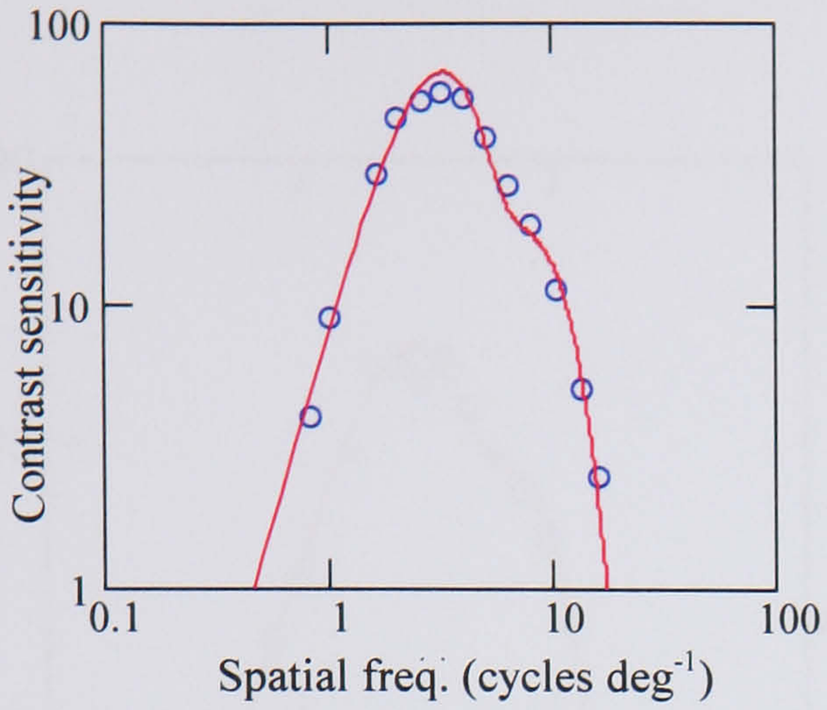


Fig. 3.8 The error plot for the sum of differentials of Gaussians model as the parameters σ_1 and σ_2 are varied and the corresponding error per data point is plotted on the surface for data set 2 Fig. 6-10 (Hawken & Parker, 1987). The contour plot of the error values is also shown.

We also did not a-priori constrain the weighting parameters and yet still found, that each curve was best described by the addition of two units, who's contrast sensitivity was lower than the maximum sensitivity of the data.

Below is shown all data from (Hawken & Parker, 1987). The first three are typical simple cells, fitted well. The parameters are ($\sigma_1, \sigma_2, a_1, a_2$). In each case the powers of differential that gave the best fit are shown in Fig. 3.9 - 3.19. Error is given as average error per data point.



$$\begin{pmatrix} 0.044 \\ 0.125 \\ 11.771 \\ 39.096 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error := 0.018

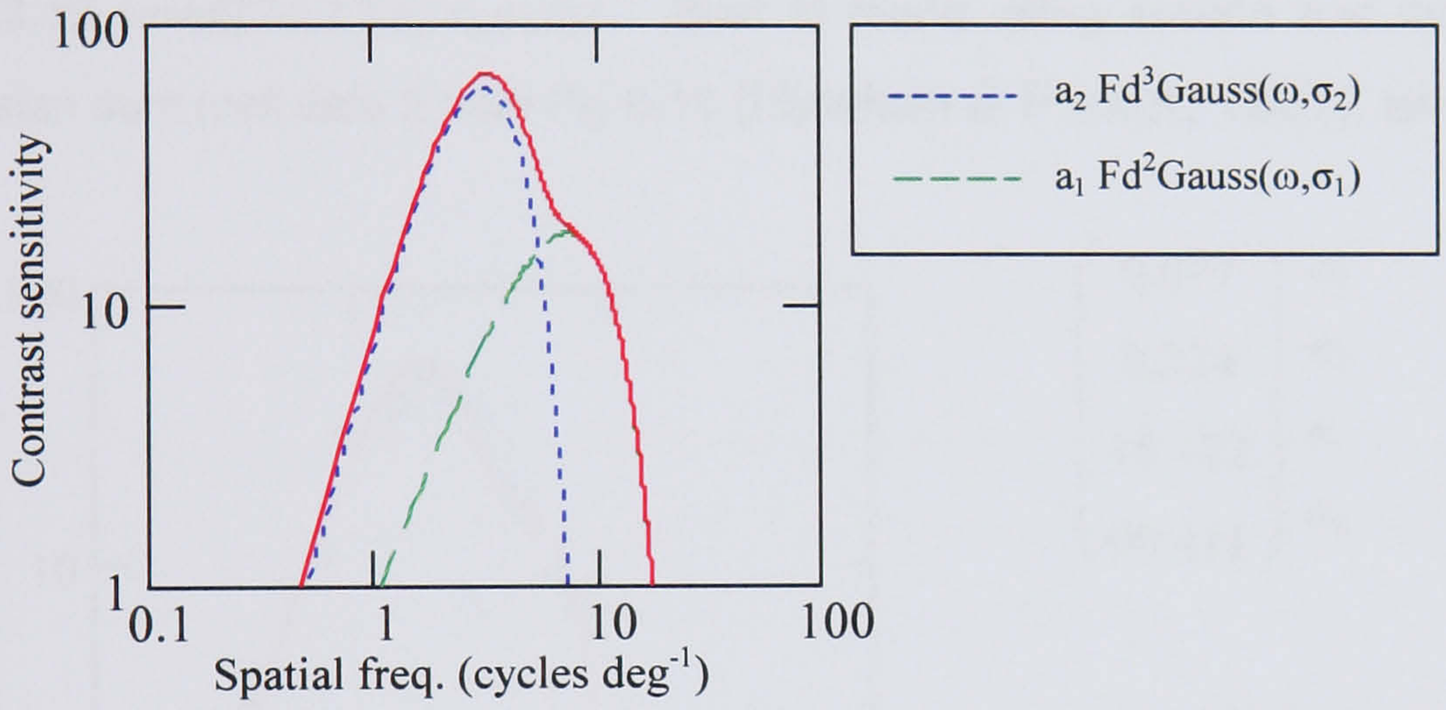
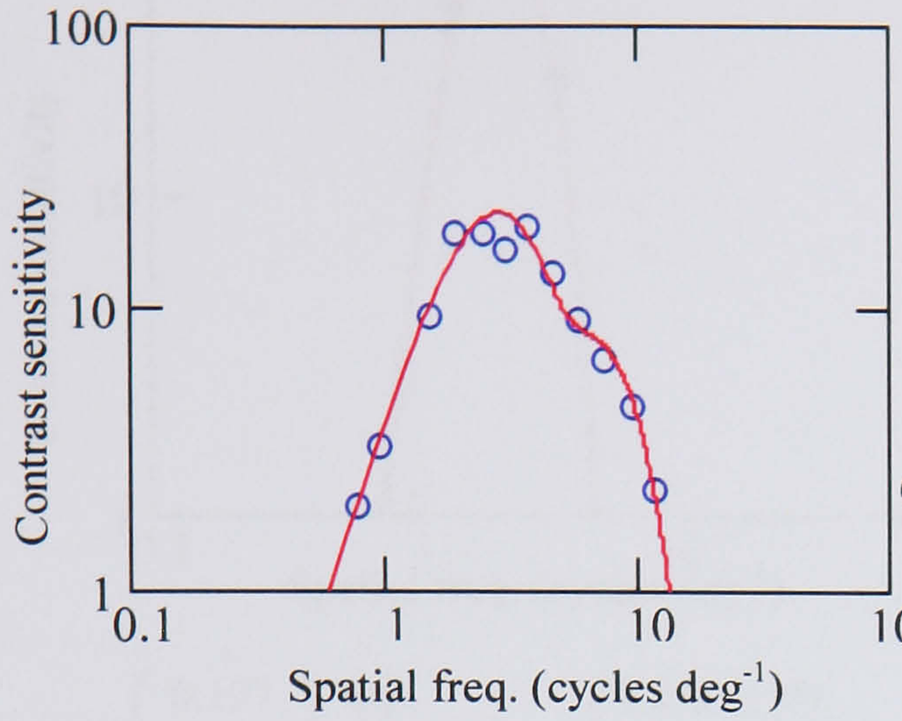


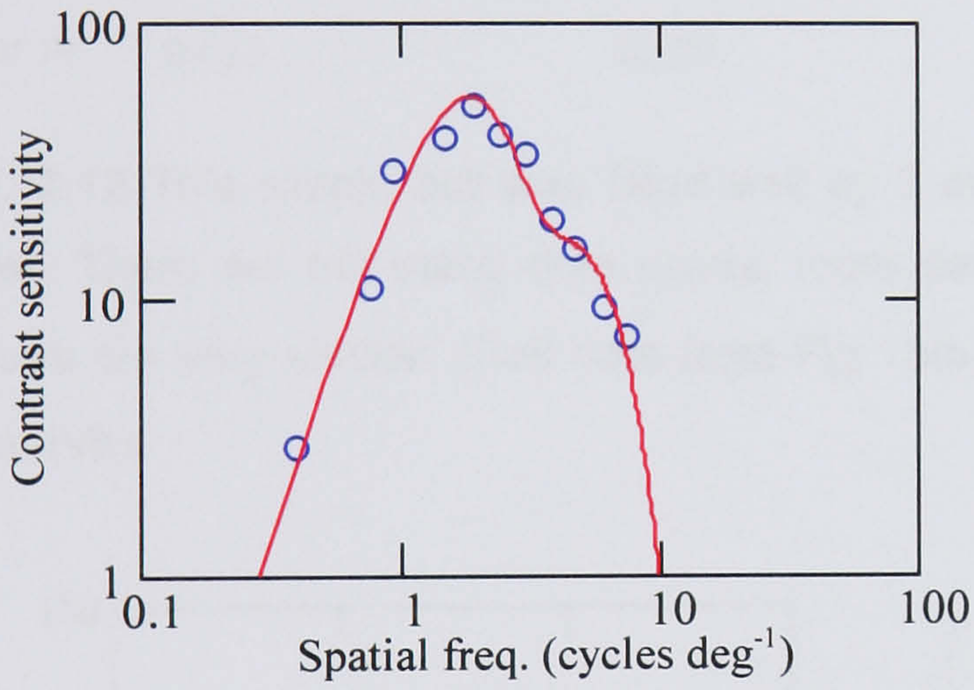
Fig. 3.9 (a) *sumdiff*$(\omega,2,3,\sigma_1,\sigma_2,a_1,a_2)$ Best fit found using second and third order differential Gaussian sum (cell data 1 from Fig 6-10 (Hawken & Parker, 1987) simple cell, layer VI) (b) shown are the two Fourier transforms of differentials of Gaussians that are the components that sum to form the function sumdiff.



$$\begin{pmatrix} 0.052 \\ 0.146 \\ 6.119 \\ 16.381 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error:= 0.016

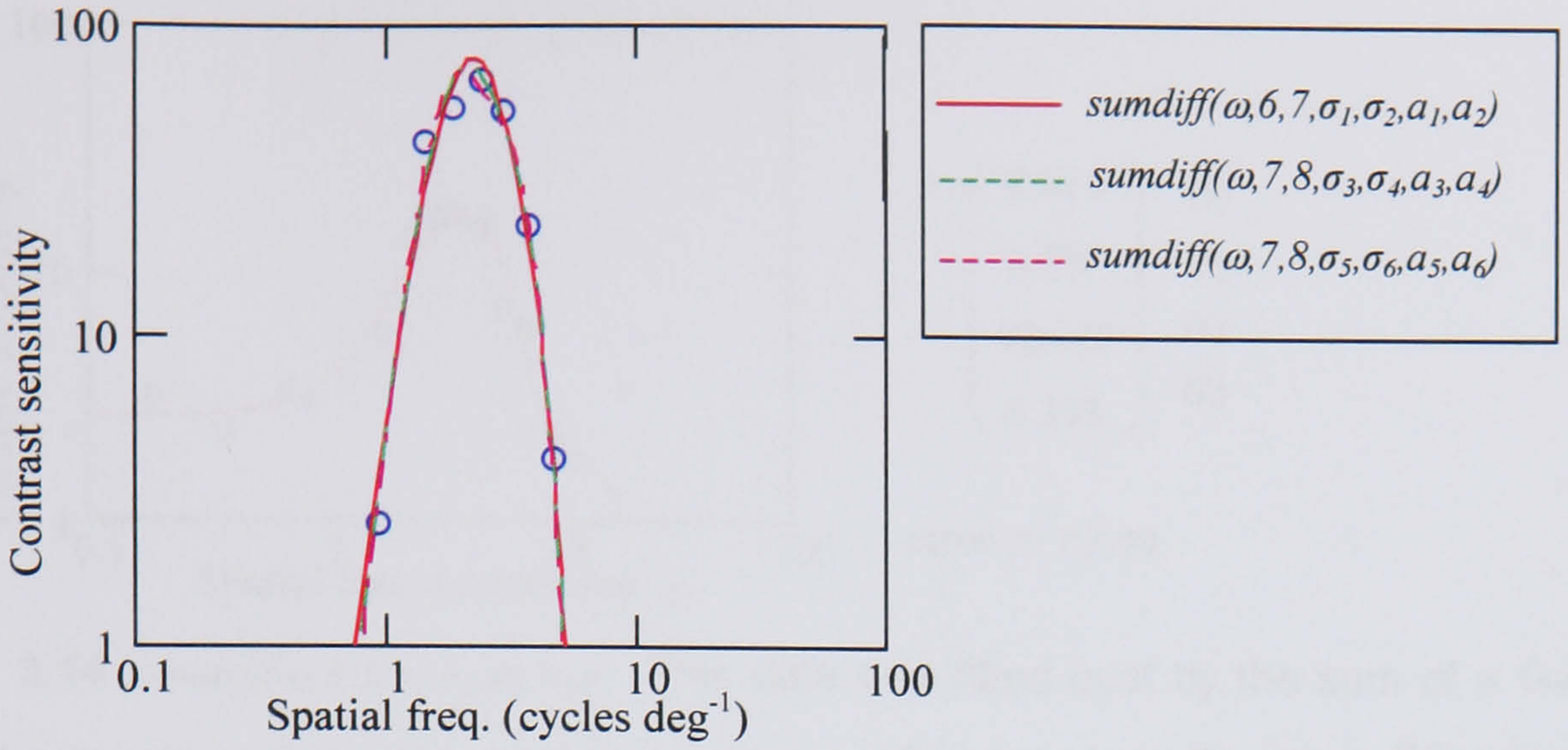
Fig. 3.10 *sumdiff*$(\omega,2,3,\sigma_1,\sigma_2,a_1,a_2)$ Best fit found using second and third order differential Gaussian sum (cell data 2 from Fig 6-10 (Hawken & Parker, 1987), simple cell, layer II).



$$\begin{pmatrix} 0.077 \\ 0.224 \\ 18.472 \\ 99.411 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error := 0.031

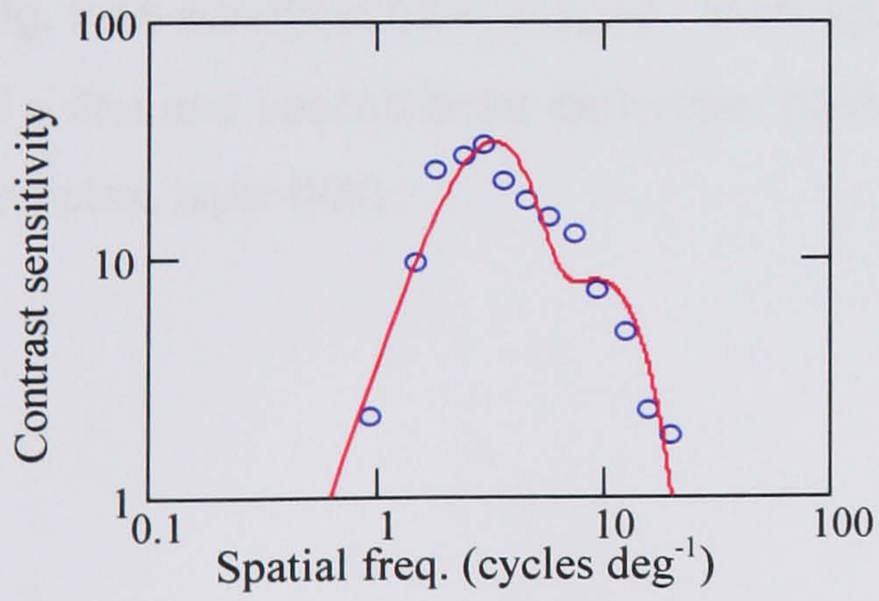
Fig. 3.11 *sumdiff*$(\omega,2,3,\sigma_1,\sigma_2,a_1,a_2)$ Best fit found using second and third order differential Gaussian sum (cell data from Fig 13(a) (Hawken & Parker, 1987), simple cell, layer II/III).



$$\begin{pmatrix} 0.197 \\ 0.286 \\ 3.551 \\ 21.742 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix} \qquad \begin{pmatrix} 0.212 \\ 0.327 \\ 2.424 \\ 24.92 \end{pmatrix} \begin{matrix} \sigma_3 \\ \sigma_4 \\ a_3 \\ a_4 \end{matrix} \qquad \begin{pmatrix} 0.224 \\ 0.364 \\ 1.36 \\ 29.823 \end{pmatrix} \begin{matrix} \sigma_5 \\ \sigma_6 \\ a_5 \\ a_6 \end{matrix}$$

error := 0.033 0.022 0.019

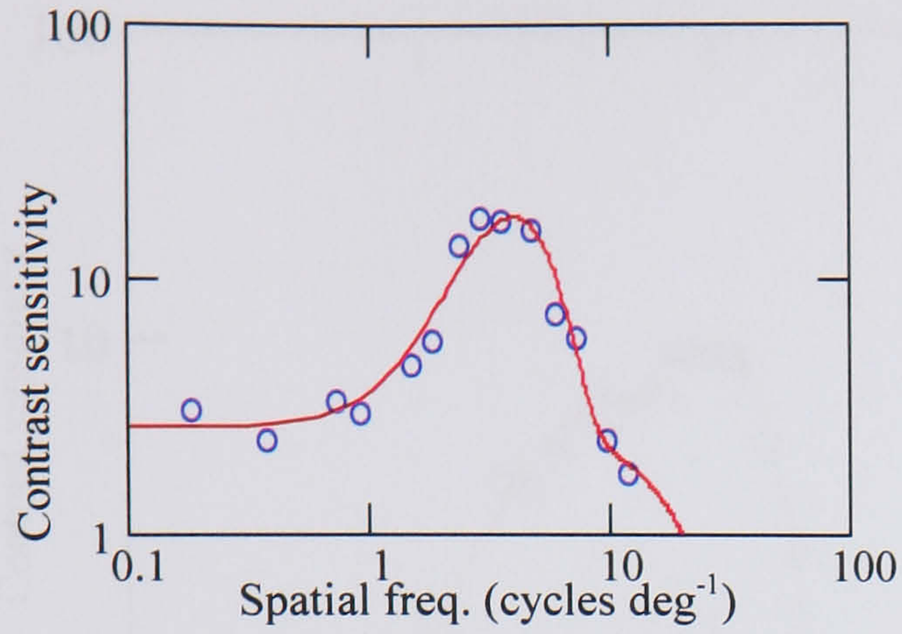
Fig. 3.12 This simple cell was fitted well by 3 different sets of adjacent powers all of a high order. There are not many data points; more data-points might resolve the fit, although all 3 curves are very similar. (Cell data from Fig 13(b) (Hawken & Parker, 1987), simple cell, layer IVb).



$$\begin{pmatrix} 0.035 \\ 0.124 \\ 4.274 \\ 18.725 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error : = 0.095

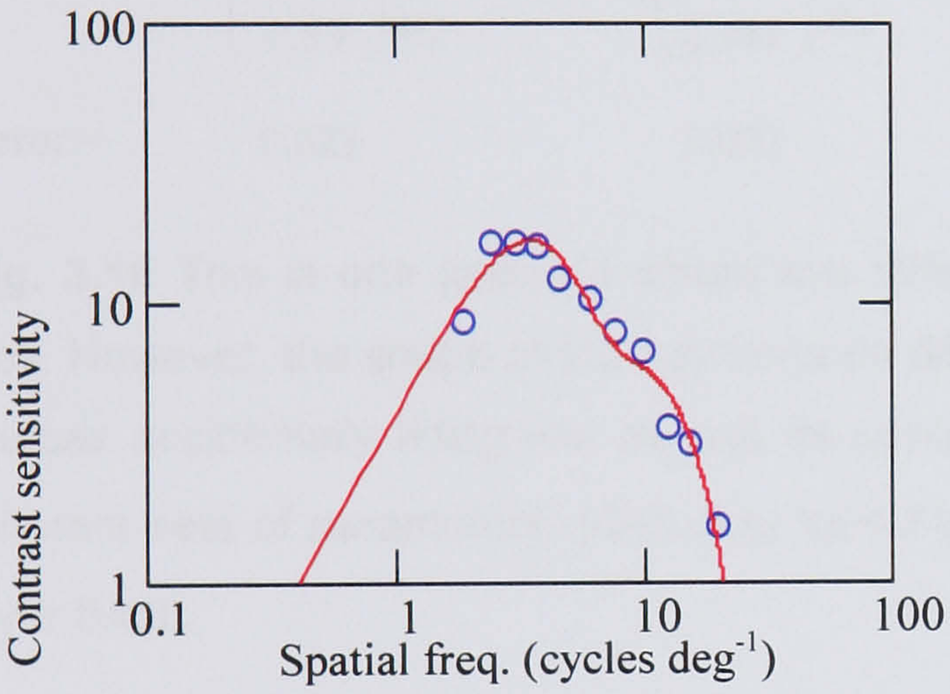
Fig. 3.13 *sumdiff*$(\omega,2,3,\sigma_1,\sigma_2,a_1,a_2)$ This is a cell the paper quotes as being complex, this is the worse fit. (Cell data from Fig 13(c) (Hawken & Parker, 1987), complex, layer V).



$$\begin{pmatrix} 0.016 \\ 0.099 \\ 92.428 \\ 6.193 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error := 0.029

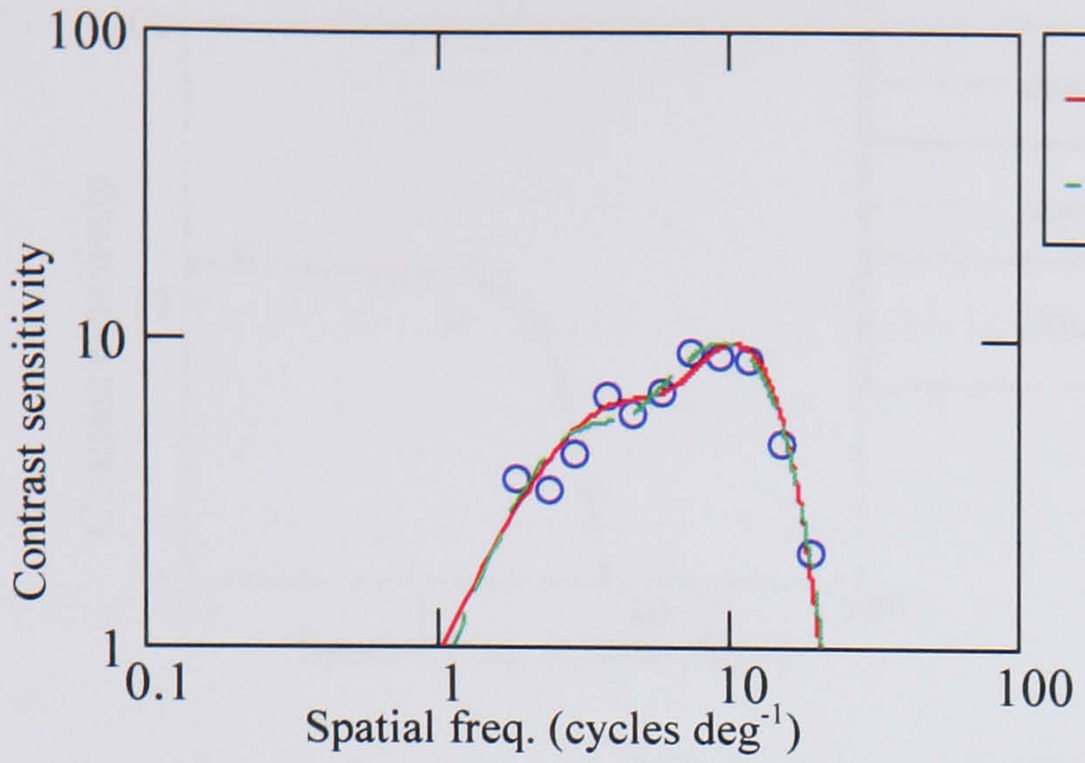
Fig. 3.14 *sumdiff*$(\omega,1,3,\sigma_1,\sigma_2,a_1,a_2)$ This data was fitted best by the sum of a Gaussian and a third order derivative, so not adjacent powers. (Cell data from Fig 13(d) (Hawken & Parker, 1987), simple, layer IVcα).



$$\begin{pmatrix} 0.028 \\ 0.098 \\ 24.72 \\ 18.846 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error := 0.021

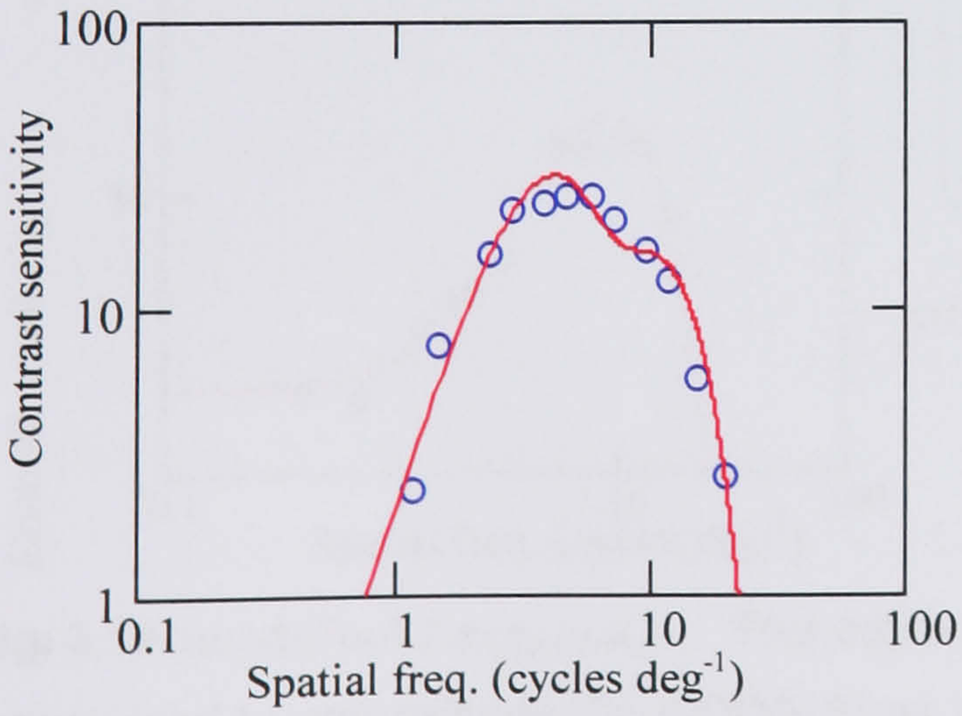
Fig. 3.15 *sumdiff*$(\omega,1,2,\sigma_1,\sigma_2,a_1,a_2)$ Here again we see a good fit, but note it is the sum this time of a first and second order derivative. (Cell data from Fig 13(e) (Hawken & Parker, 1987), complex, layer II/III).



$$\begin{pmatrix} 0.094 \\ 0.038 \\ 6.162 \\ 0.59 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix} \qquad \begin{pmatrix} 0.034 \\ 0.145 \\ 4.964 \\ 2.441 \end{pmatrix} \begin{matrix} \sigma_3 \\ \sigma_4 \\ a_3 \\ a_4 \end{matrix}$$

error:= 0.023 0.022

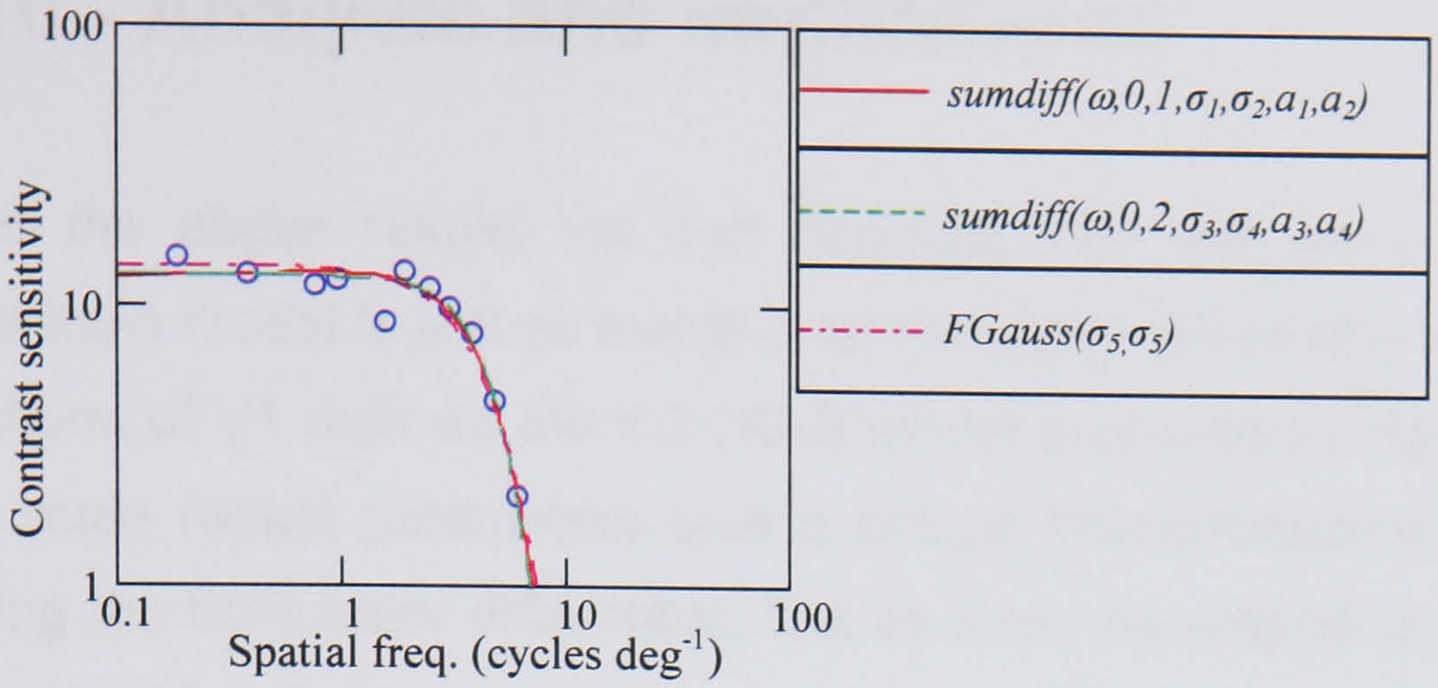
Fig. 3.16 This is one example where two different sets of parameters fitted the data equally well. However, the shape of the functions do differ slightly, making this more a case of different shapes accidentally fitting just as well as opposed to the same curve being described by two different sets of parameters. (Cell data from Fig 13(f) (Hawken & Parker, 1987), complex, layer IVcβ).



$$\begin{pmatrix} 0.034 \\ 0.102 \\ 7.954 \\ 10.043 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error := 0.02

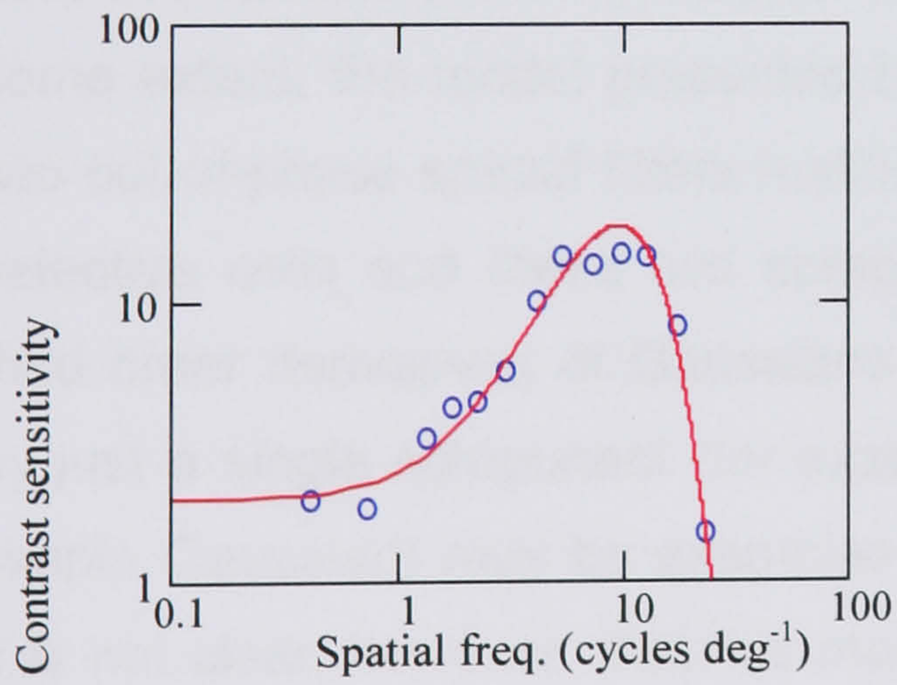
Fig. 3.17 *sumdiff(ω,2,3,σ1,σ2,a1,a2)* Best fit found using second and third order differential Gaussian sum (Cell data from Fig 13(g) (Hawken & Parker, 1987), complex, layer VI).



$$\begin{pmatrix} 0.077 \\ 0.087 \\ 90.984 \\ 8.189 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix} \qquad \begin{pmatrix} 0.084 \\ 0.098 \\ 85.736 \\ 4.228 \end{pmatrix} \begin{matrix} \sigma_3 \\ \sigma_4 \\ a_3 \\ a_4 \end{matrix} \qquad \begin{pmatrix} 0.072 \\ 108.044 \end{pmatrix} \begin{matrix} \sigma_5 \\ a_5 \end{matrix}$$

error:= 0.018 0.015 0.021

Fig. 3.18 This set of data could be described well by the Fourier transform of a Gaussian, but slightly better by adding a small scale first or second order Gaussian. (Cell data from Fig 13(h) (Hawken & Parker, 1987), simple, layer VI).



$$\begin{pmatrix} 0.014 \\ 0.034 \\ 78.66 \\ 8.681 \end{pmatrix} \begin{matrix} \sigma_1 \\ \sigma_2 \\ a_1 \\ a_2 \end{matrix}$$

error : 0.023

Fig. 3.19 *sumdiff*$(\omega,0,2,\sigma_1,\sigma_2,a_1,a_2)$ This cell wasn't fitted as well by the simple DOG model (a model used to approximate the d-DOG-S) as by the Gabor, but this function manages to fit it well. However, the two powers are not adjacent. (Cell data from Fig 14 (Hawken & Parker, 1987)).

3.10 - Analysis and implications

From the above results we can conclude that the sums of derivatives of Gaussians model is just as suitable for modelling the spatial contrast sensitivity functions of V1 cells as the d-DOG-S model proposed by Hawken and Parker. The more robust parameters give a unique characterisation of each cell. By adding the third order differential, it is as if we are extending the second order differential function to make it more accurate. We can consider cutting down the number of parameters by defining the powers as n and $n+1$, which would leave us with 5 parameters, however this will lead in two of the cases to much worse fits. We have also seen that as with the case of the d-DOG-S function the new function fits complex and simple cells alike.

In (De Valois et al., 2000), principal components analysis was used to find the main components of the space-time oriented receptive fields of directionally selective simple cells. They found two main spatial components in quadrature with each other. They suggested that the directional cells would receive these multiple linear inputs from non-directional cells. Their study is restricted to simple cells and we do not have information of the direction selectivity of the cells in (Hawken & Parker, 1987), however we do find close fits to the data. To some extent, the model presented here is supporting evidence for the idea of two out-of-phase spatial filters making up the main components of directionally selective cells and these two components are modelled well by second and third order derivatives of Gaussians. Furthermore, cells that were fitted better by just a single component (for example where we used just a transform of a simple Gaussian) may be examples of earlier non-directionally selective cells. It is not clear, however, that this model involving the linear combination of two components would apply to complex cells as they combine input in a non-linear fashion.

What does it mean to find good fits for clearly complex cells as well? Most complex cells are directionally selective (Hubel & Wiesel, 1962), so if the model

presented here applies to directionally selective cells, than maybe we should not be surprised that it fits the complex data well. However the model is the linear combination of linear cells, which means the output should have linear properties, which is not the case for complex cells. This matter needs further investigation.

Curve fitting was also carried out with a simple non-linear combination of these sub-units, adding them and then squaring them. This was expected to converge to a better fit as all parameters produce a positive value. In general though, this function did not fit any better than the simple weighted sum, although it mostly gave very similar results in terms of error and gave first and second order derivative combinations as best fits. Allowing the weightings to be negative, i.e. to fit the square of the difference; sometimes gave slightly better results than the initial model, but introduced a degeneracy of parameters in which several functions with very different parameters could fit equally well.

3.11 - Conclusion

The d-DOG-S model was presented as a superior model for a primary spatial filter, using fitted contrast sensitivity functions as evidence in the paper of Hawken and Parker (1987). It was suggested that these models provide a detailed description of the organisation of the sub-regions of the receptive field according to the parameters found for each fit. However the number of free parameters was discarded as not important, when in fact, the very number used makes the curve fitting procedure a pointless exercise, caused by the over-definition of the curve (Bevington, 1992; Stark, 1997). Furthermore it is erroneous to infer that contrast sensitivity function fits give information on the spatial receptive field for both simple and complex cells when we are not sure of their linear properties. We also managed to show that these curves might

have redundant interchangeable parameters, which means that the set of parameters given for each curve has no specific meaning for each cell.

An alternative point of view presented here is that it is not necessary to use a model that involves as many parameters as d-DOG-S to produce similarly accurate fits. This was achieved with models that make use of the blurring-differentiating effect of derivatives of Gaussian functions. We found a simpler model, with less parameters, that converged onto a unique set of parameters for each cell and in general fitted just as well. We demonstrated that it is not only the difference of Gaussians model that can be extended to provide more accurate fits than previous simpler models. We extended the differential of Gaussians model. As well as fitting this data well, the sum of derivatives of Gaussians model must necessarily fit well all previous data that was well described by the second differential of a Gaussian, as this is just a special case of the model, with one of the space constants = 0. It is possible that cells that are well fitted by the simpler version are earlier along in the visual pathway and the cells shown here are higher in the hierarchy of V1, combining the inputs of the single unit cells. This model suggests that cell responses to contrast over spatial fequency in V1 are described by the sum of the Fourier transforms of the differentials of Gaussians. As the different orders of derivatives can be taken to represent different spatial filtering channels, this fit ties in with the notion that directionally selective simple cells in V1 are linear combinations of two different non-directionally selective components. However as the model presented here fits complex cells equally well, this theory could potentially be extended to directionally selective complex cells as well.

The models that will be introduced and investigated in this thesis will be models of spatial representation and motion calculation based on the notion that this is achieved from the combined input of derivative of Gaussian shaped responses of V1 cells. Here we have shown that the Fourier transform of these types of functions can be successfully used to describe the contrast sensitivity functions of cells in V1, though previously it had been suggested that they could not.

Therefore it is biologically plausible to construct models based on these mathematical functions. In the past, the derivatives of Gaussians approach has often been supported for strong theoretical signal processing based reasons (Koenderink & van Doorn, 1987; Young, 1985). This chapter has added to the body of physiological evidence that supports the use of derivatives of Gaussians based models to describe V1 behaviour.

First of all, the possibility of multiplying different derivatives was investigated as a possible model as derivative based methods for motion calculation involve multiplication. However, this did not appear to describe the contrast sensitivity functions of the cells presented here. It is possible that the addition is a step towards performing a multiplication, e.g. by making use of an identity such as: $a \cdot b = \frac{(a+b)^2 - (a-b)^2}{4}$. However it is not clear why the visual system would implement multiplication in such an apparently inefficient way. Multiplication can also be achieved by additive integration and non-linear transfer characteristics in synapses (Hildreth & Koch, 1987).

In the Multi-Channel Gradient model a Taylor series based representation of the scene is used to calculate motion (Johnston et al., 1999) (see Chapter 4). This representation, before it is used to calculate motion involves adding together differently weighted differentials, which are calculated using derivatives of Gaussians filters. The cells modelled in this chapter may be subunits of this spatial representation. In the next chapter we describe such a Taylor series based representation in more detail and how it is used to calculate motion. The following modelling work is based on the assumption that the derivatives of Gaussians all have the same space constant. It is the different size of space constants that produced the unusual shapes of the curves in this chapter, but adding two curves of the same size constant is a subset of these functions and this makes calculation simpler. We can then consider the possibility of combining different scale derivatives.

Chapter 4- Taylor series based spatial representation and the McGM motion model

I have shown that linear combinations of differentials of Gaussians can be used to successfully fit the contrast response of simple and complex cells even in cases where it had been suggested that an approach based on differentials of Gaussians would not provide a good model of cell responses. Therefore it is biologically plausible to propose filters that take the form of derivatives of Gaussians as building blocks of various visual algorithms. In this chapter we will consider how these functions in the spatial domain can be used to build an information rich representation of the scene, illustrating this process with reconstructions of images. I will then describe how these spatial representations can also be formed across time using log Gaussian temporal filters. I will explain in detail how the Multi Channel Gradient Model of motion uses these filters to construct an output from a sequence of images containing the calculated image velocity. Finally, in this chapter I adapt the Multi Channel Gradient Model to form a reconstruction of the input sequence from its spatio-temporal representation.

4.1 - Taylor series based spatial representation

The following theory is based on the idea of representing local geometry using so called 'local jets', that is truncated Taylor series expansions (Koenderink & van Doorn, 1987). We have studied cell responses in the striate cortex, and functions were suggested that could fit these responses. We now examine the

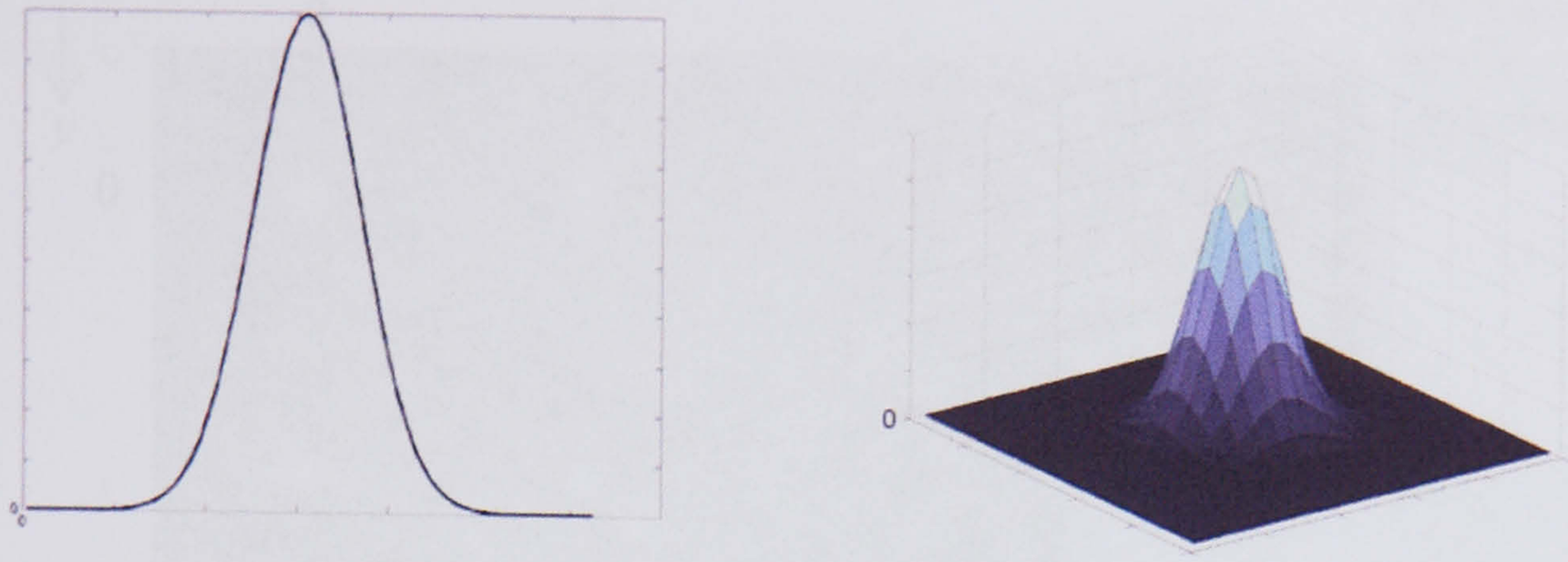
computational use of such functions. The local jet will be used to combine cell responses in to a useful representation of local image brightness.

Below is the equation for the two dimensional Taylor series expansion about a point *(x,y)* of the function *f(x,y)*.

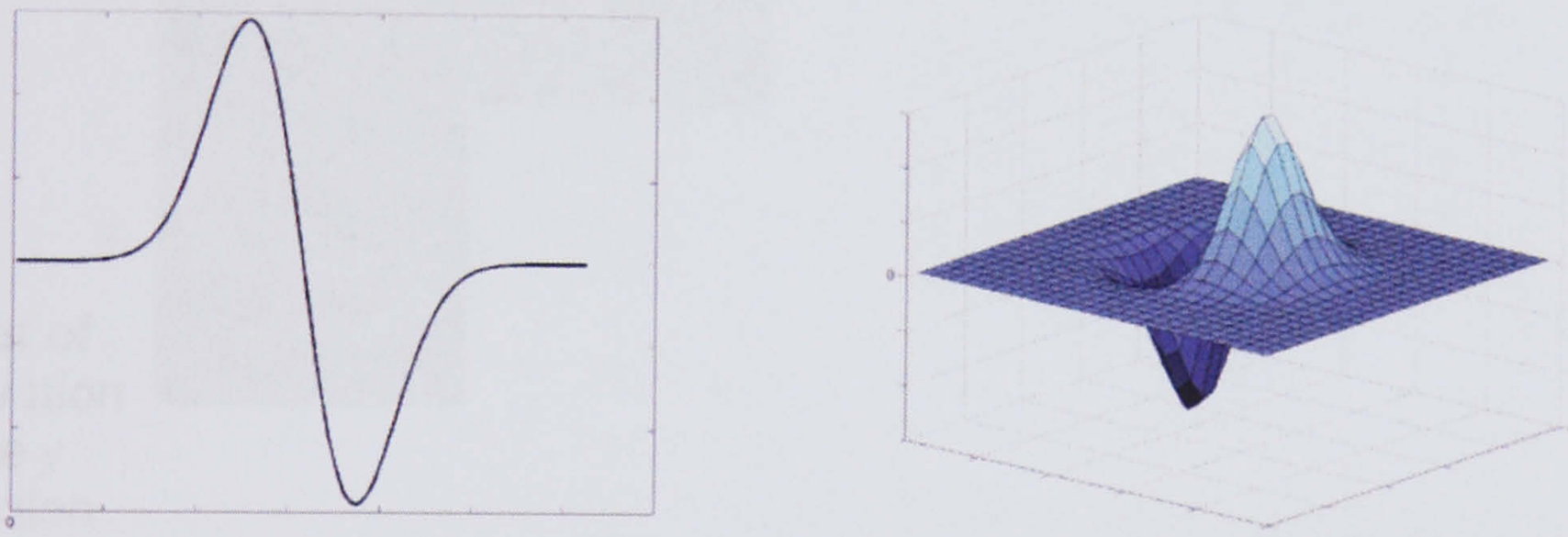
$$f(x+p,y+q)=f(x,y)+p\frac{\partial f(x,y)}{\partial x}+q\frac{\partial f(x,y)}{\partial y}+\frac{1}{2!}\left(p^2\frac{\partial^2 f(x,y)}{\partial x^2}+q^2\frac{\partial^2 f(x,y)}{\partial y^2}+2pq\frac{\partial^2 f(x,y)}{\partial xy}\right)+... \tag{4.1}$$

This implies that within a given neighbourhood of the point *(x,y)* in an image one can approximately predict the brightness of nearby points given that one knows the value of enough derivatives of image brightness at the point *(x,y)* to be suitably accurate. This holds true if it is possible to define the derivatives of the image brightness at all points, i.e. *f(x,y)* is a continuous function. Smoothness can be ensured by blurring (i.e. *smoothing*) the image first. The image is convolved with a two dimensional Gaussian function.

a



b



c

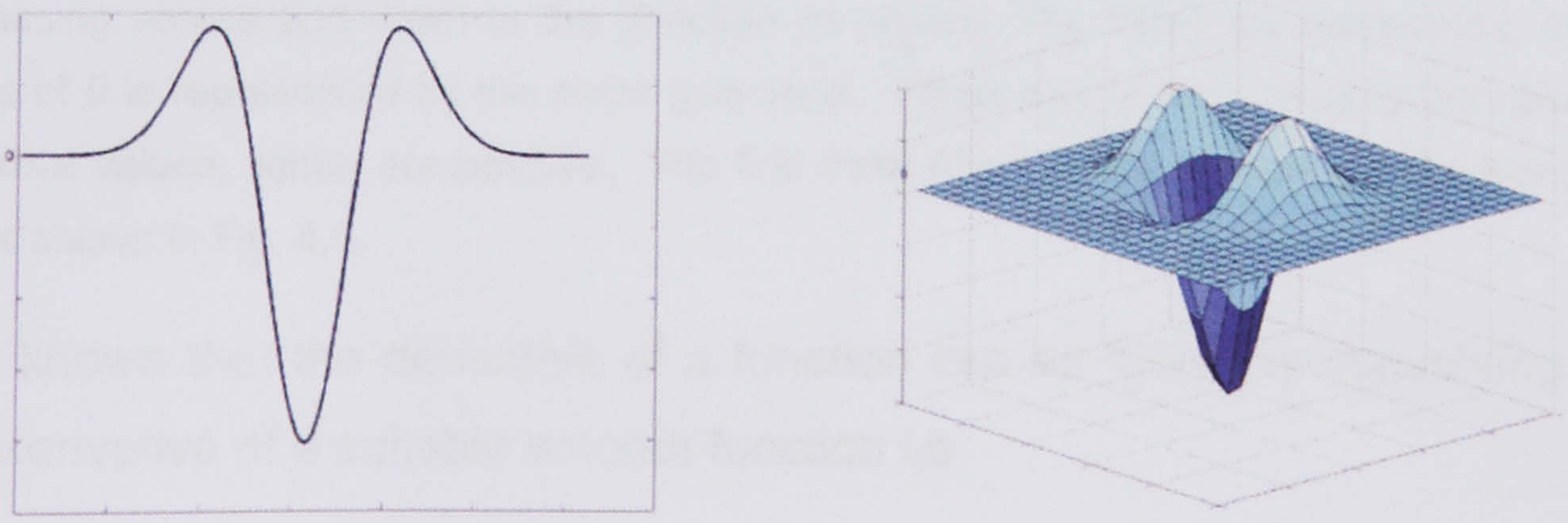


Fig. 4.1 Plots of 1D and 2D Gaussian derivatives in the x direction. (a) 0 order, (b) 1st order (c) 2nd order

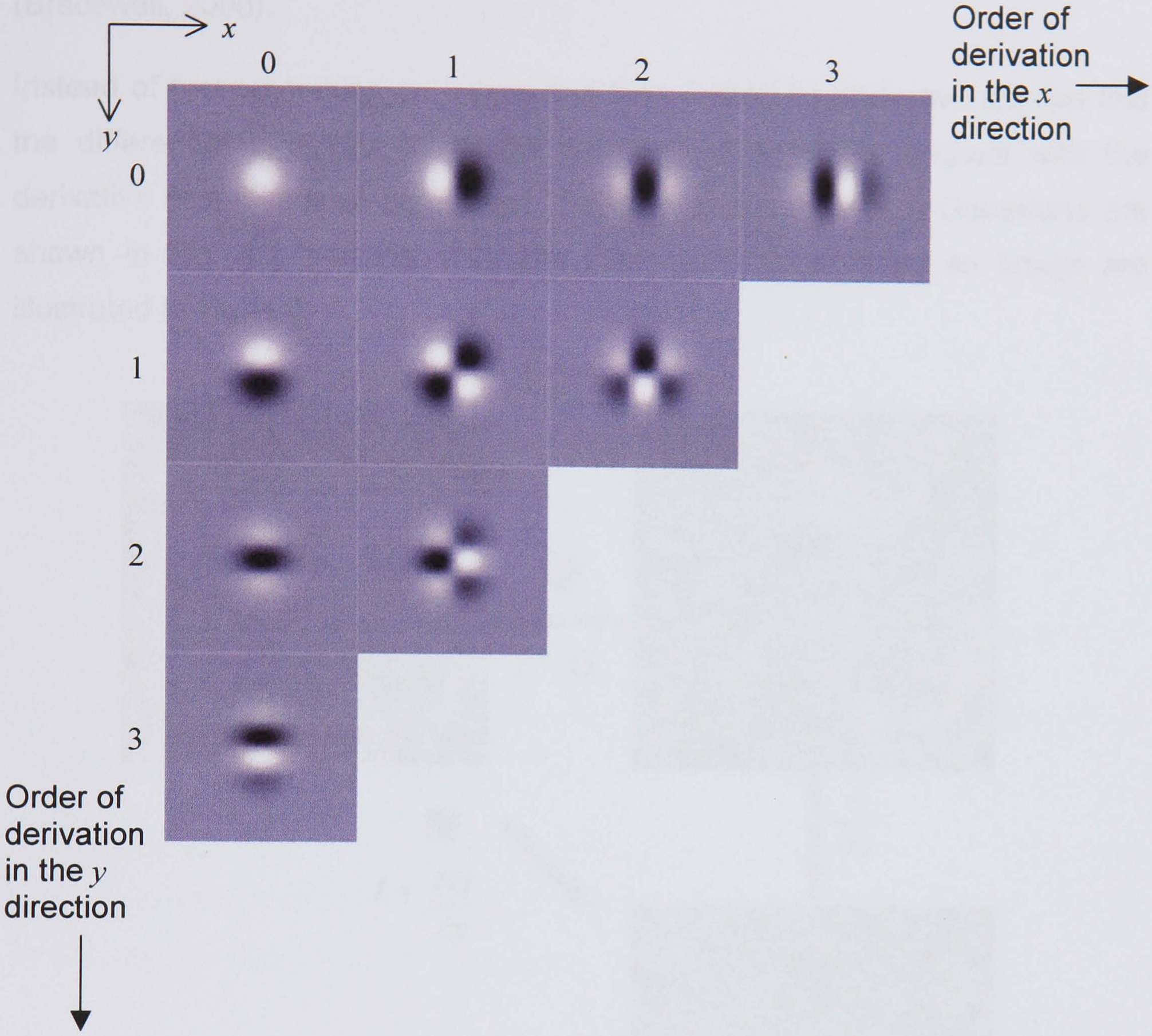


Fig. 4.2 Plots of Gaussian partial derivative filter kernels in the x and y directions, with order increasing across and down in the direction as shown. The filters are normalised, so that a value of 0 is represented by the same grey level. Pixels darker then the grey background are negative values, lighter are positive. The first three filters in the top row are the same as the filters shown in Fig. 4.1.

It is known that the derivative of a function can be found by convolving it with the derivative of a suitable smooth function i.e.

$$D\big(G(x,y) * f(x,y)\big) = DG(x,y) * f(x,y) \tag{4.2}$$

Where D is the differential operator and G is the 2D Gaussian function (Bracewell, 2000).

Instead of first smoothing an image and then finding its derivative we can find the differential of a blurred image via convolution of the original with the derivative of a Gaussian filter (Marr, 1982) – 2D Derivatives of Gaussians are shown in Fig. 4.1 and Fig. 4.2. The steps for differentiating an image are illustrated in Fig. 4.3.

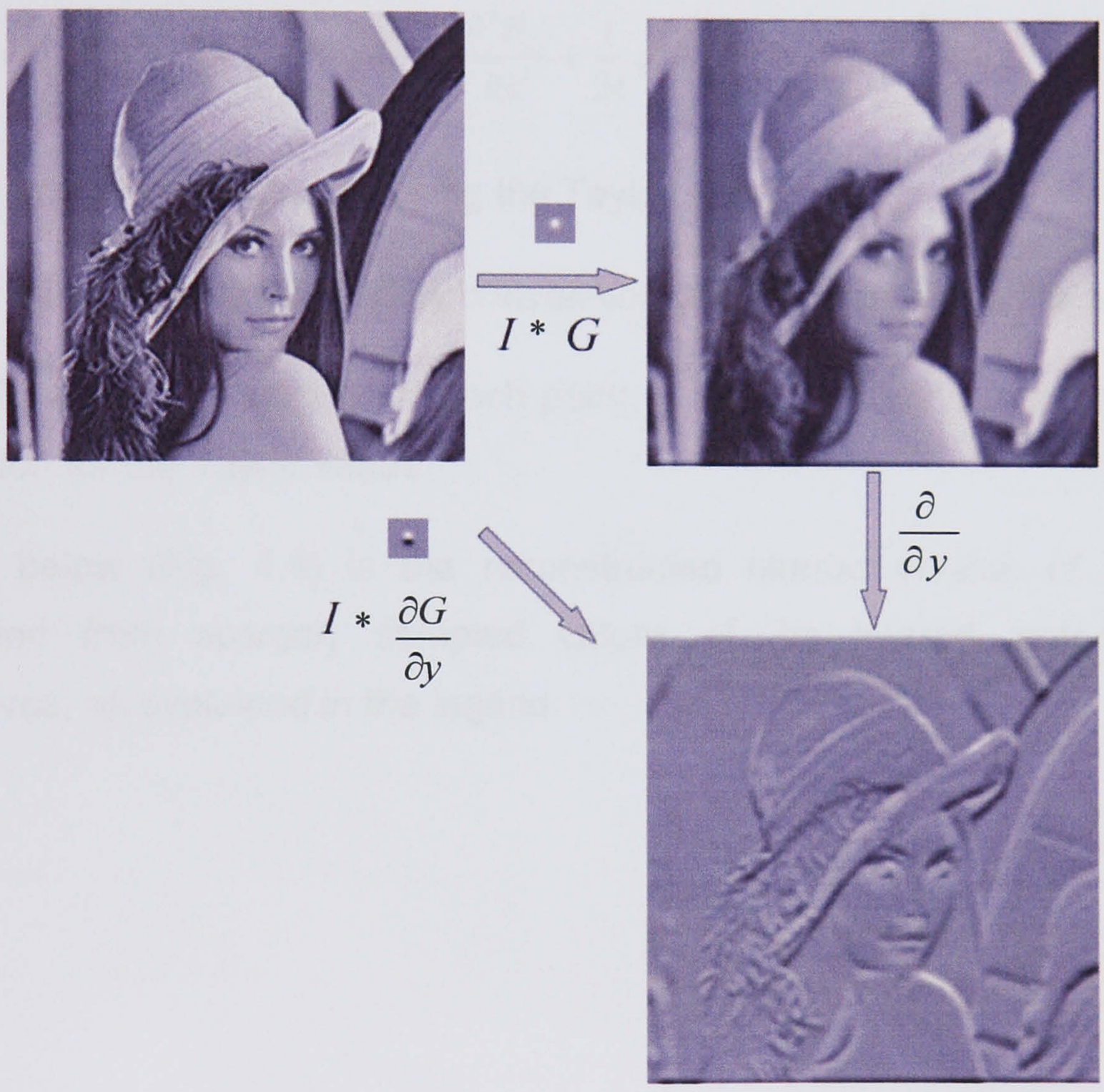


Fig. 4.3 The image is convolved with the kernel $\partial G/\partial y$, where $G(x,y)$ is the Gaussian function. This is the same as convolving the image with G and finding its derivative. Note that convolving with this kernel results in an output with luminance changes in the y direction emphasised.

A useful feature of the Gaussian filter is its separability so that

$D_{x,y}(G(x,y)) = D_x G(x) * D_y G(y)$, where D_x denotes partial differentiation w.r.t. x and G is the Gaussian function. Hence, a 2D convolution can be calculated from two 1D convolutions in series. So in fact only the 1D spatial derivatives of Gaussians up to maximum order need to be calculated to form the components of the Taylor expansion.

In this way we can use the different filtered outputs in the Taylor series approximation.

$$R_{i+p,j+q} = B_{i,j} + p\frac{\partial B_{i,j}}{\partial x} + q\frac{\partial B_{i,j}}{\partial y} + \frac{1}{2!}p^2\frac{\partial^2 B_{i,j}}{\partial x^2} + \frac{1}{2!}q^2\frac{\partial^2 B_{i,j}}{\partial y^2} + \frac{2}{2!}pq\frac{\partial^2 B_{i,j}}{\partial xy}... \quad (4.3)$$

R is the image reconstructed using the Taylor series

B is the blurred image formed by convolution of the image with a Gaussian

The values p, q are weights at each point, given by distance from the point of expansion for the Taylor series.

Shown below (Fig. 4.4) is the reconstructed blurred version of an image generated from sparsely sampled values of the blurred image and its derivatives, as explained in the legend.

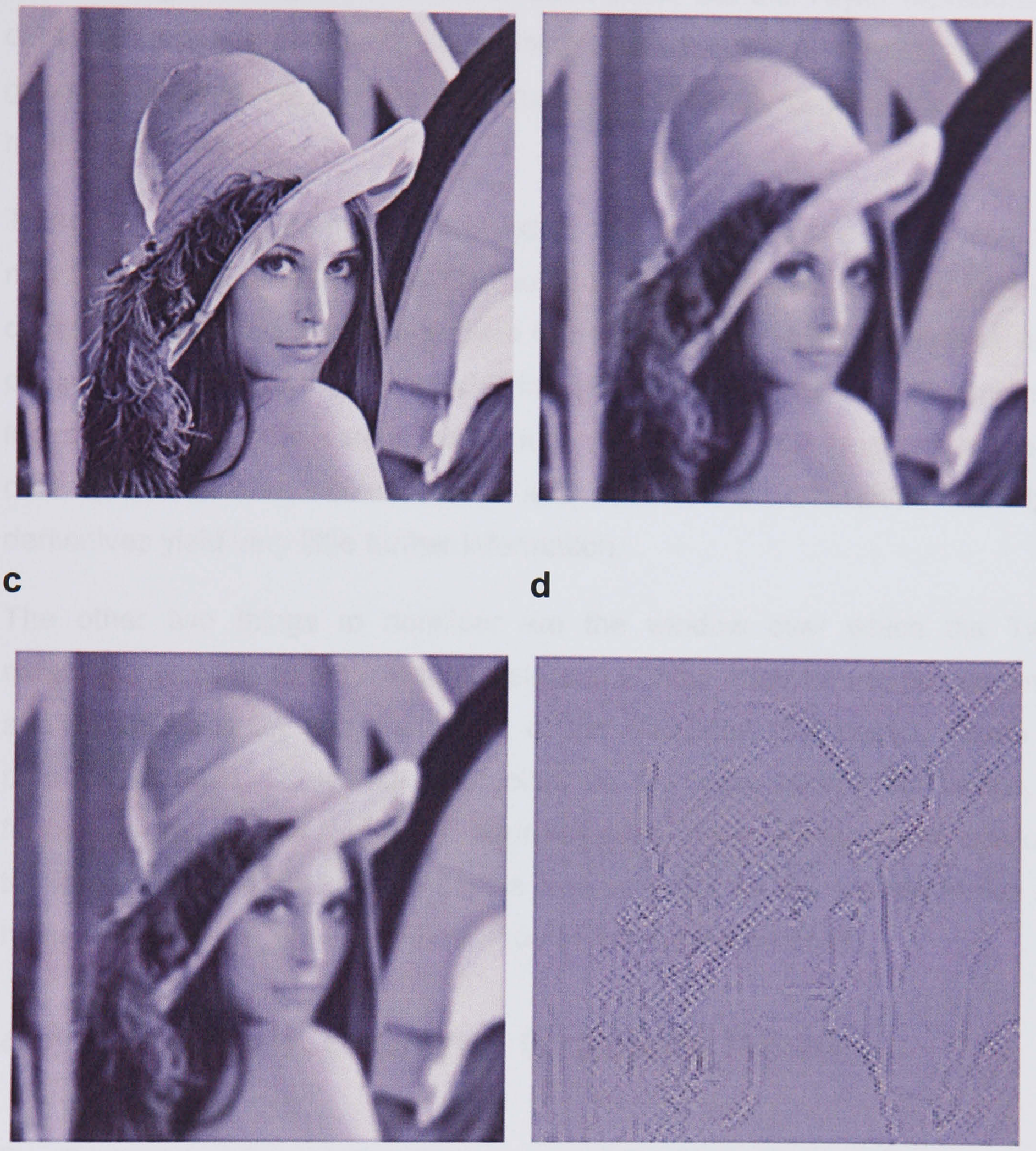


Fig. 4.4 Reconstructing an image using Gaussian derivatives. (a) The original image, 256×256 pixels. (b) The image blurred with a 23×23 pixel Gaussian kernel, σ = 1.5. (c) The image reconstructed in the way described above, using derivatives of the same Gaussian kernel of up to the 3rd order. Taylor expansions extended over 3×3 windows, so every 9th pixel is sampled from each convolution output. (d) The scaled difference between the blurred and reconstructed image, max = 0.7% of max image brightness (white) and min = -0.8% (black).

The Taylor representation in this form is not a more efficient coding than representing the brightness value at each point, but the Taylor representation contains more information, describing the image at several layers of geometrical structure in a way that can be used for further calculations such as motion extraction.

There are a few parameters involved in this reconstruction. In our choice of maximum order of differential filter to use, we are limited to some extent by computational time, and in any case the calculation of high derivatives is not plausible as typically receptive fields have a limited number excitatory/inhibitory lobes (De Valois & De Valois, 1990; Hubel & Wiesel, 1962). Using up to the fifth order provides reasonable results, and after the 5^{th} differential the higher derivatives yield very little further information.

The other two things to consider are the window over which the Taylor expansion is used to approximate values (i.e. how often to sample the image and its differentials) and the extent of the Gaussian blur kernel. These two parameters trade off against each other as the more blurred the image, the further we can extend the Taylor approximation. If the window of expansion is too large, the reconstruction becomes less accurate, as the approximation only holds true within a given neighbourhood of the expansion point.

4.2 - The Multi Channel Gradient Model

The Taylor jet representation is the basis of the Multi Channel Gradient Model of motion extraction, which will now be described in more detail. Above, the two dimensional Taylor series in x and y was described, with x and y referring to horizontal and vertical directions across an image respectively. In order to calculate motion the model also makes use of the Taylor series at different orientations, where the directions x_θ, y_θ remain orthogonal, but θ is chosen along different orientations. The number of orientations to choose is one of the parameters of the model. Typically 24 values for directions are calculated. The

final velocity calculation uses all of these directions. Oriented filters can be constructed from filters that are differentials in the x and y directions using appropriate weights. This steering is implemented in the model (Dale, 2003). Oriented filters represent the orientation tuned response within orientation columns in V1 (Albright et al., 1984; Hubel & Wiesel, 1974). The model uses up to 5th order derivatives in the direction of the filter orientation and up to 2nd order derivatives in the orthogonal direction.

In order to extract the magnitude of the change in space over time we also need to calculate differentials involving time. In order to represent image structure over time the Taylor series needs to incorporate the outputs of temporal filters. These are modelled by log Gaussians and their derivatives (up to the second derivative) filters, that match the shape of the sensitivity profiles from transient channels in the visual system derived from psychophysical measurements (R. F. Hess & Snowden, 1992; Johnston & Clifford, 1995). These filters have two parameters associated with them, τ (the standard deviation) and α (the delay) which are explained in Fig. 4.5. The output of these filters combines information over several frames of motion input. The Taylor series representation is now constructed as a function of three variables x, y and t. As above, the log Gaussian is separable and the convolution with temporal filters can be performed in sequence before the spatial convolution.

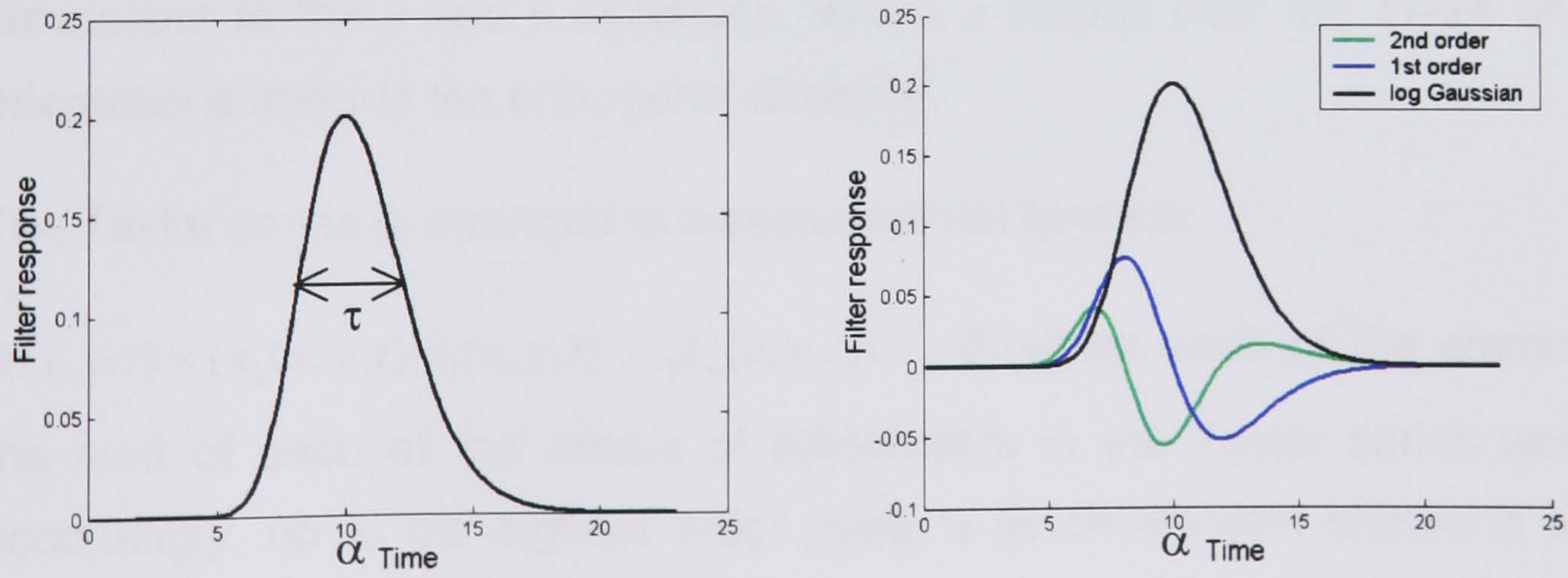


Fig. 4.5 The log Gaussian and its derivatives. α = 10, the delay from onset to peak of the response of the log Gaussian. τ = 0.275, the standard deviation of the response in log time.

Once the bank of differential filters have been constructed we can use them to construct the Taylor series representation as explained above in each of the directions. Now we are using a three dimensional Taylor series representation in two directions in space and in time.

$$f(x+p,y+q,t+r) =$$

$$f(x,y,t)+\left[p\frac{\partial f(x,y,t)}{\partial x}+q\frac{\partial f(x,y,t)}{\partial y}+r\frac{\partial f(x,y,t)}{\partial t}\right]+\ldots$$

$$+\frac{1}{2!}\left[p^2\frac{\partial^2 f(x,y,t)}{\partial x^2}+q^2\frac{\partial^2 f(x,y,t)}{\partial y^2}+r^2\frac{\partial^2 f(x,y,t)}{\partial t^2}+\ldots\right.$$

$$\left.+2pq\frac{\partial^2 f(x,y,t)}{\partial xy}+2pr\frac{\partial^2 f(x,y,t)}{\partial xt}+2qr\frac{\partial^2 f(x,y,t)}{\partial yt}\right]+\ldots \qquad \textbf{(4.4)}$$

The Taylor series components are calculated by first finding the temporal derivatives by convolving the sequence with the derivatives of log Gaussians (up to the second order) and then convolving each order of the temporally blurred image with the spatial filters in turn, which are two dimensional Gaussians in the x and y directions, where x ranges over the m=24 different orientations and y is the orthogonal direction.

The Taylor series is arranged in a vector-valued function

$\boldsymbol{k}(x,y,t) = (k_0(x,y,t),k_1(x,y,t),\ldots,k_n(x,y,t))^{\mathrm{T}}$, in which each of the elements is the sum of each of the orders of differentials in the Taylor series weighted accordingly, up to the highest order used, n (each element shown in square brackets in Eqn. 4.4). The sums that form the elements of the vector valued function are represented in Chapter 3 in a single dimension and used to fit the V1 cells' contrast sensitivity function.

The first step of the motion calculation is to take the derivative of this vector valued function, resulting in a matrix that in turn also has elements comprised of sums of differentials of the image. In practise we do not need to perform this initial step but calculate the derivatives and associated weights as needed for the matrix

$J = \mathrm{D}\,\boldsymbol{k}(x,y,t)$ - the derivative of the vector valued function $\boldsymbol{k}(x, y, t)$

$= (\boldsymbol{k}_x(x,y,t), \boldsymbol{k}_y(x,y,t), \boldsymbol{k}_t(x,y,t))$ - where $\boldsymbol{k}_x(x,y,t)$ is the derivative w.r.t. x of $\boldsymbol{k}(x, y, t)$

$$= \begin{bmatrix} k_{0,x} & k_{0,y} & k_{0,t} \\ k_{1,x} & k_{1,y} & k_{1,t} \\ \vdots & \vdots & \vdots \\ k_{n,x} & k_{n,y} & k_{n,t} \end{bmatrix}, \quad (4.5)$$

where $k_{i,x}$ = partial derivative w.r.t. x of the ith element of the vector valued function $\boldsymbol{k}(x, y, t)$, i.e. the sum of the ith order derivatives of the Taylor expansion. We can now compute the matrix product

$$J^{\mathrm{T}}J = \begin{bmatrix} \boldsymbol{k}_x \cdot \boldsymbol{k}_x & \boldsymbol{k}_x \cdot \boldsymbol{k}_y & \boldsymbol{k}_x \cdot \boldsymbol{k}_t \\ \boldsymbol{k}_y \cdot \boldsymbol{k}_x & \boldsymbol{k}_y \cdot \boldsymbol{k}_y & \boldsymbol{k}_y \cdot \boldsymbol{k}_t \\ \boldsymbol{k}_t \cdot \boldsymbol{k}_x & \boldsymbol{k}_t \cdot \boldsymbol{k}_y & \boldsymbol{k}_t \cdot \boldsymbol{k}_t \end{bmatrix}. \quad (4.6)$$

Where $\boldsymbol{k}_x \cdot \boldsymbol{k}_x = k_{0,x} \cdot k_{0,x} + k_{1,x} \cdot k_{1,x} + \ldots + k_{n,x} \cdot k_{n,x}$, the dot product of $\boldsymbol{k}_x$ with itself.

This matrix is integrated over a spatiotemporal volume $R = a \geq p \geq b, c \geq q \geq d, e \geq r \geq f$ to give the matrix

$$M = \int_e^f \int_c^d \int_a^b J^{\mathrm{T}}J \; \mathrm{d}p\mathrm{d}q\mathrm{d}r = \begin{bmatrix} \boldsymbol{x} \cdot \boldsymbol{x} & \boldsymbol{x} \cdot \boldsymbol{y} & \boldsymbol{x} \cdot \boldsymbol{t} \\ \boldsymbol{y} \cdot \boldsymbol{x} & \boldsymbol{y} \cdot \boldsymbol{y} & \boldsymbol{y} \cdot \boldsymbol{t} \\ \boldsymbol{t} \cdot \boldsymbol{x} & \boldsymbol{t} \cdot \boldsymbol{y} & \boldsymbol{t} \cdot \boldsymbol{t} \end{bmatrix}. \quad (4.7)$$

By integrating over the partial derivatives we arrive at the values of the vectors x, y and t over the spatial temporal volume.

In practise the integration is achieved by summation over each of the variables (detailed in (Johnston & Clifford, 1995)). This matrix has terms x, y, t, that each are constructed from sums of terms of the basic Taylor image representation partially differentiated w.r.t. x, y and t respectively. From this matrix we can use some of the terms to recover well-conditioned estimates of image velocity in two orthogonal spatial directions, namely $x \cdot t / x \cdot x$ and $y \cdot t / y \cdot y$. The denominator is equal to the squared magnitude of a vector, e.g. $|x|^2$. This scalar product is only zero when all the terms of the vector are zero, i.e. when the image is uniform, in which case image velocity is undefined. However, raw speed measures such as $x \cdot t / x \cdot x$, are still infinite for directions parallel to isobrightness contours.

The calculations shown above are calculated for m = 24 different orientations of the x direction. Using all the results we have produced so far in all the different orientations we can now define speed, $\hat{s} = (\hat{s}_{\parallel}, \hat{s}_{\perp})$, a vector whose components are speed and orthogonal speed. Similarly we can construct inverse speed $\check{s} = (\check{s}_{\parallel}, \check{s}_{\perp})$.

$$\hat{s}(\theta) = \sqrt{\frac{2}{m}} \left[\frac{x_\theta \cdot t_\theta}{x_\theta \cdot x_\theta} \left(1 + \left(\frac{x_\theta \cdot y_\theta}{x_\theta \cdot x_\theta} \right)^2 \right)^{-1}, \frac{y_\theta \cdot t_\theta}{y_\theta \cdot y_\theta} \left(1 + \left(\frac{x_\theta \cdot y_\theta}{y_\theta \cdot y_\theta} \right)^2 \right)^{-1} \right] \tag{4.8}$$

$$\check{s}(\theta) = \sqrt{\frac{2}{m}} \left(\frac{x_\theta \cdot t_\theta}{t_\theta \cdot t_\theta}, \frac{y_\theta \cdot t_\theta}{t_\theta \cdot t_\theta} \right) \tag{4.9}$$

Each of these constitute a $m \times 2$ matrix, where m is the number of orientations used for the motion calculation (m = 24) and $\sqrt{\frac{2}{m}}$ is a normalization factor. (This step is further explained in Johnston et al., 1999). With these two matrices we can now combine speed measures over different orientations to calculate image motion over the input frames. This is done by projecting onto fiducial

sine and cosine functions, for which we construct normalized cosine and sine vectors

$$\boldsymbol{F}(\theta) = (F_{\parallel}(\theta), F_{\perp}(\theta)) = \sqrt{2/m}\left(cos(\theta), \sin(\theta)\right). \tag{4.10}$$

Using this matrix as a fiducial reference frame, in terms of θ for the calculation of motion direction, we can extract the fundamental Fourier coefficients of the directional speed functions. Final overall speed squared is computed as a ratio of determinants:

$$S^2 = \frac{\begin{vmatrix} \hat{s}_{\parallel} \cdot F_{\parallel} & \hat{s}_{\parallel} \cdot F_{\perp} \\ \hat{s}_{\perp} \cdot F_{\parallel} & \hat{s}_{\perp} \cdot F_{\perp} \end{vmatrix}}{\begin{vmatrix} \hat{s}_{\parallel} \cdot \check{s}_{\parallel} & \hat{s}_{\parallel} \cdot \check{s}_{\perp} \\ \hat{s}_{\perp} \cdot \check{s}_{\parallel} & \hat{s}_{\perp} \cdot \check{s}_{\perp} \end{vmatrix}}. \tag{4.11}$$

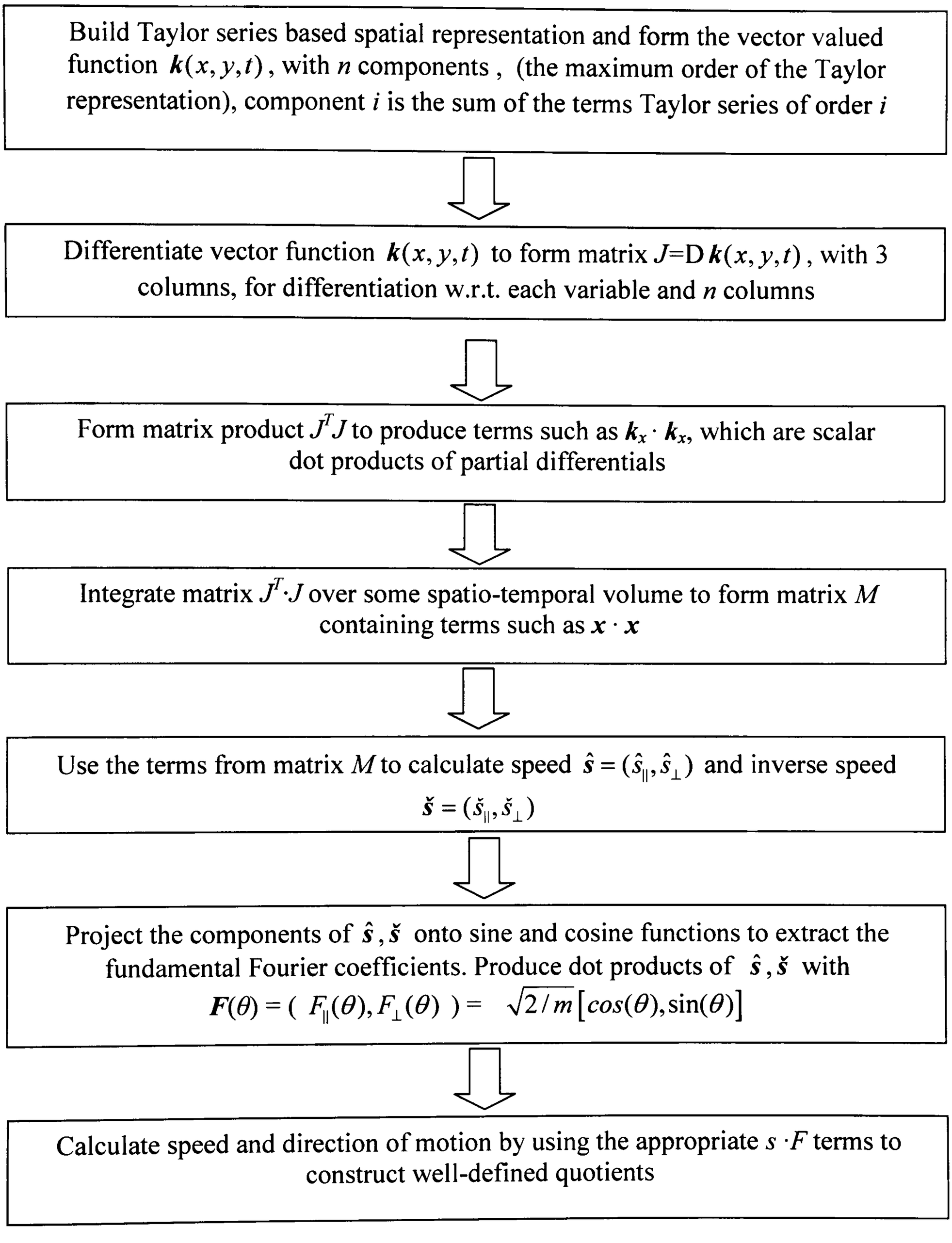
Where $\hat{s}_{\parallel} \cdot F_{\parallel} = \sum_{\theta} \hat{s}_{\parallel}(\theta) \cdot F_{\parallel}(\theta) = \sqrt{2/m} \sum_{\theta} \hat{s}_{\parallel}(\theta) \cdot \cos(\theta)$.

The direction can also be computed explicitly using these terms:

$$\text{direction} = \tan^{-1}\left(\frac{\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\parallel} - \left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\perp}}{\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\perp} + \left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\parallel}}\right) \tag{4.12}$$

Where $\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right)$ indicates the pair-wise multiplication of the elements of the two vectors $\check{s}_{\parallel}, \hat{s}_{\parallel}$ to form a vector whose terms are $\left(\check{s}_{\parallel}(\theta) \times \hat{s}_{\parallel}(\theta)\right)$. In (Johnston et al., 1999), instead of multiplication $\left(\check{s}_{\parallel} + \hat{s}_{\parallel}\right)$ is used. In practice there is little difference between the two methods in terms of the output of the motion model.

This is considered to be the phase angle of the projection on fiducial sine and cosine functions as encoded by pairs of cells in the visual system. These are the main steps of calculation in the model, see below for a hierarchical step-by-step overview.



For the motion calculation, further parameters have been introduced on top of the ones described for the spatial reconstruction. For the temporal blur we have τ, the size of the blur of the log Gaussian and α, the delay between the onset of the log Gaussian and its peak (as shown above). The number of temporal

derivatives is fixed at 2. The next parameter is the number of orientations used ($m = 24$), and finally we must fix the size of the integration zone (11 pixels).

With these parameters the McGM model has been shown to accurately detect motion in real life scenes as shown in Fig. 4.6. (Dale, 2003). It has also been successful in describing the perceived speed and direction of second order motion (Johnston & Clifford, 1995).

Fig. 4.6 A typical output from the motion for a real input scene of traffic motion. The model picks out the rightward motion of the car and the leftward motion of the pedestrian. Model parameters: spatial blur $\sigma = 1.5$, spatial blur support area= 23×23 pixels, temporal filter parameters: $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Top left: one frame from the input sequence. Top right: one frame from the temporally blurred sequence. Bottom left: Velocity magnitude (brighter areas indicate greater velocity). Bottom right: Motion direction (direction indicated by the colourwheel).

4.3 - Applying the motion model

Demonstrated below is how the model performs the motion calculation with typical stimuli that I will be attempting to model in the context of a motion – spatial position interaction. Before constructing a combined model we need to

see how the separate filter responses in the spatial representation and the motion calculation behave. The model outputs a sequence that is the original sequence blurred in space and time, and a corresponding sequence of frames of motion calculations for velocity magnitude and direction. For all future examples of output generated by the model, the input consists of a sequence of 256×256 pixels images.

First, let's take a single flash presented for only one frame as an input stimulus. This is not the most straightforward input for motion calculation as this sequence contains large transients. However, an impulse in time is useful for examining the temporal response of the filters.

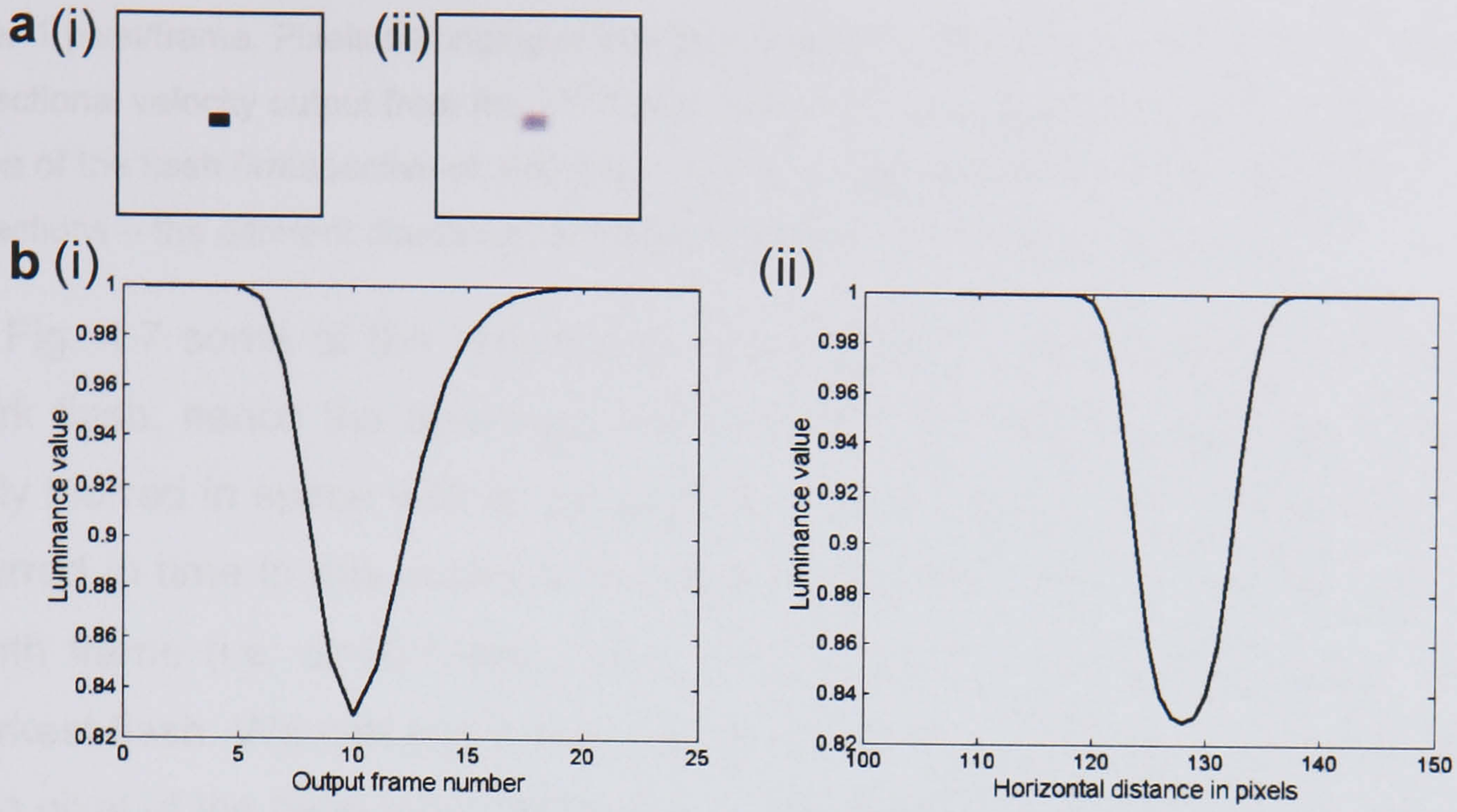


Fig. 4.7 The results of a sequence containing a single frame flash. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. (a) (i)The area of the input frame containing the flash. All of the images in the input the sequence are blank apart from the one containing the 5×9 black flash in the middle. (ii) The output frame in which the spatio-temporally blurred output shows the most presence of a flash, the flash is blurred spatially and appears in several output frames. (b) (i) The luminance values at one pixel in the flash over the frames from the spatio-temporally blurred output. The response peaks 10 frames after the presentation of the flash. (ii) The spatial luminance profile horizontally in line with the flash in the frame shown in (a)(ii). The hard edged bar luminance is blurred in a Gaussian shape. (c)(i) The output from the 11^{th} frame of the velocity calculation (the flash is presented in frame 0). Thresholded for values

over 1 pixel/frame. Pixels belonging to the flash location lie within the black outline. (ii) The directional velocity output from the 11th frame. (d) (i) The average velocity magnitude over the area of the flash (irrespective of direction). (ii) The magnitude of velocity averaged over directions – the different directions cancel each other out, leaving no overall velocity.

In Fig. 4.7 some of the results from the model are shown (note that this is a dark flash, hence the downward impulse function). We first look at the output only blurred in space with a Gaussian and in time with a log Gaussian. As it is blurred in time in this output a faint flash appears in several frames, but in the tenth frame (i.e. α=10 frames after the presentation of the flash) we find the darkest flash. We can see this in Fig. 4.7(b). The output luminance over time at one pixel of the flash reflects the shape of the temporal filter. Spatially, the hard edges of the flash are blurred out in a Gaussian shape. Taking the eleventh frame from the velocity output we can see that the flash is calculated to have an associated velocity in all directions, which spreads out beyond the perimeter of the flash (Fig. 4.7(c)). If we take the average velocity at each frame, this fluctuates dramatically, although in 10th frame, where flash representation is strongest there is no associated velocity. However, if we average over the directions, because all directions are present there is no overall velocity in a single direction.

We'll now take a look at the motion calculation for a drifting grating.

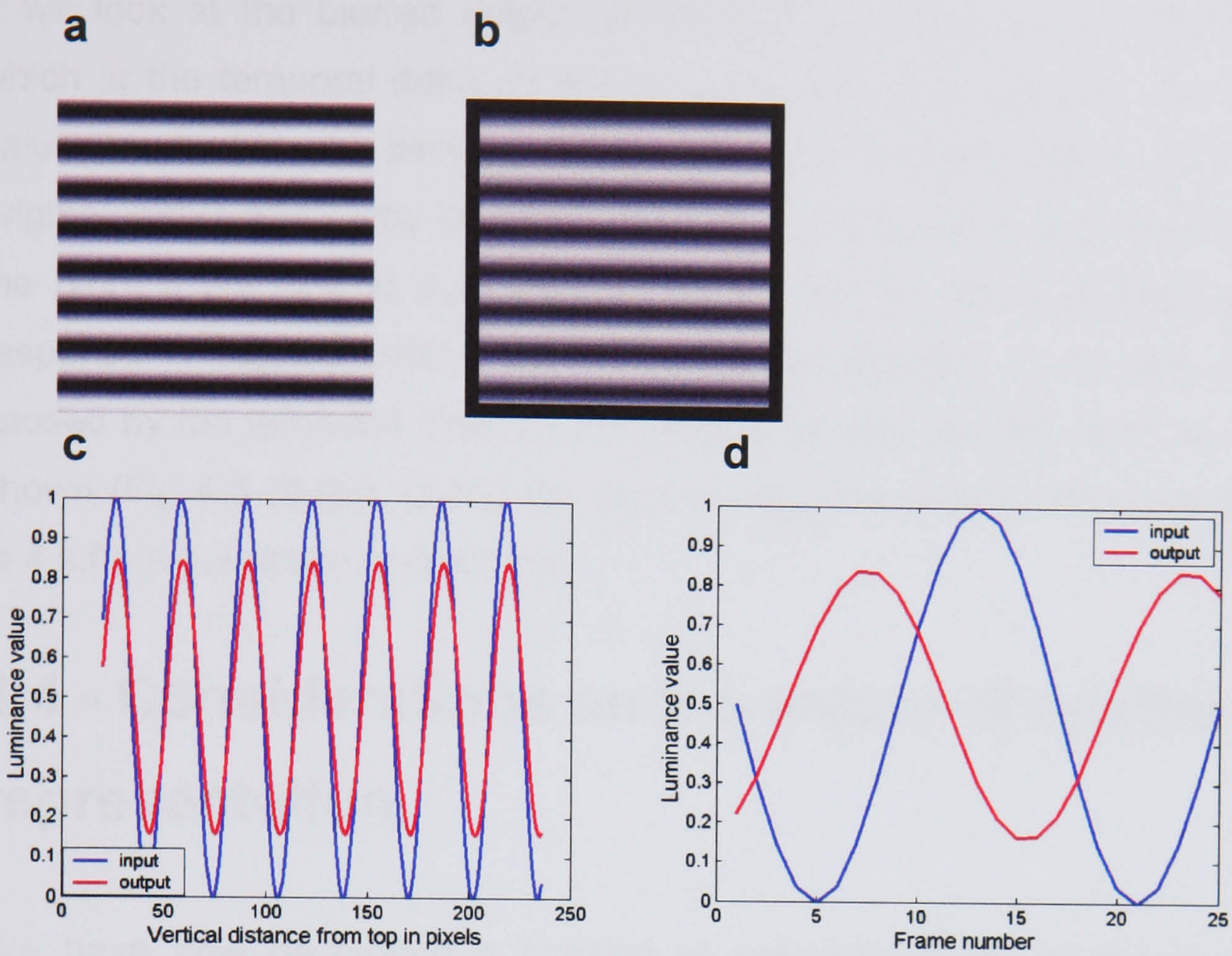


Fig. 4.8 Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. (a) Input: an upwardly drifting sine wave grating (2 pixels/frame). (b) The 10th frame from the output that is the spatio-temporal blur. (c) The luminance profile of the grating in a vertical line from the top (8 cycles/image) (blue), plotted along with the luminance profile of the blurred output from 10 frames later (red). (d) The luminance of the middle pixel of the input over each frame (1/16 cycles/frame) plotted against the luminance of the middle pixel of the blurred output over time. (e) 10th frame from the velocity magnitude output. (f) 10th frame from the velocity direction output.

If we look at the blurred output corresponding to the input 10 frames later, which is the temporal delay of the temporal filter (Fig. 4.8 (c)), we see lower values caused by the blurring and we see that it is the nearly in phase to the original, although slightly lagging behind (the sine curve is moving to the left on the plot). If we look at a single pixel over time, we can see that the blurred response is lagging behind by about (but not exactly) 10 frames, the delay caused by the temporal filter. All the velocity outputs are the same as the ones shown (Fig 4.8 (f),(g)), giving the correct response of 2 pixels/frame (accurate to 4 s.f.) in the upwards direction.

4.4 - Considerations on the nature of spatial representation

We have now developed a method of representing an image by applying derivatives of Gaussian filters to produce derivatives of the blurred image. These are then summed with appropriate weights to reconstruct the blurred image using the Taylor approximation. We can use this technique not only to represent a single image, but also a sequence of images. We also know that we can use the Taylor representations of a sequence of images to extract the motion field from a sequence at different points in time using the Multi Channel Gradient Model (Johnston et al., 1992; Johnston & Clifford, 1995; Johnston et al., 1999). I am going to extend this motion model to use the differentials of the blurred images taken from the sequence to build an output sequence of blurred image representations. I will test the results of the model on sequences of moving images (of 256×256 pixels). We are considering the Taylor series representation as formed in V1 and the motion analysis to be taking place in MT+ areas. The input sequence is taken to be the actual stimulus as presented in the real world. V1 cells can be characterised as linear filters constructed from responses from the retina through the LGN to result in their particular receptive field shapes.

Using the Taylor reconstruction we can only represent the blurred (using a given blur kernel and its derivatives) version of an image. We are going to use the same blurred differential images as used to calculate motion. The original blur kernel applied by the McGM results in a sequence with elements blurred in time using a 1D log Gaussian kernel as well as in space using a 2D Gaussian kernel. We begin by only reconstructing the spatial aspect of the visual scenes using the 2D Taylor series in the x and y directions. (See "Further work" Chapter 8 section for ideas on multi-orientation reconstruction). We have seen in the example of the drifting grating that blurring in time means that we cannot directly link the blurred sequence elements in time with each of the input frames as none of the frames from the blurred sequence match the phase of a given frame from the input exactly. In this sense the processing does not proceed 'a frame at a time' and although we can consider the framerate of the input and we can calculate an associated velocity accordingly, the rebuilt sequence operates on a different timescale and can be considered the basis of the perceptual report. In this way, as has been suggested in the past (Rao et al., 2000; Johnston & Nishida, 2001) we dissociate between 'real time' and 'perceptual time'.

It is important to consider what this Taylor series representation means in the physical world. Crucially, it is not simply a reconstruction of the input sequence blurred in space and time, in fact it is not suggested that the original image is ever reconstructed in the brain – although this might eventually be necessary. Rather, the Taylor series is a meaningful representation that exists when considering the summed activity of V1 cells weighted according to the Taylor weights. Such a representation provides a rich description of the visual scene that can be used for many functions not just motion calculation. V1 has been cited as performing many important tasks rather than just being an accurate topographic map of the scene. For example V1 has been implicated in processes such as grouping, segmentation and contour matching (Gallant et al., 1995; Kapadia et al., 1995).

We are considering each window over which a local Taylor series expansion is calculated as a V1 aggregate receptive field. These windows however, are necessarily proportional to the size of the spatial blur applied by the motion model as an accurate reconstruction is only possible if the reconstruction window is small enough in comparison with the spatial blur (see above). With the reconstruction algorithm as it stands we assume a constant receptive field size. This means we do not take into account the fact that receptive field size increases with eccentricity in V1. This is one of the simplifications in the model, which will limit us in applying it to investigate gradual effects of eccentricity, but greatly simplify the calculations. Another simplification is that these windows have a perfectly square structure, which of course is not the case in any biological system. In this case it is the topology of these receptive fields that we are concerned with and it will become apparent that this is no barrier to performing the function of representing all manners of real stimulus. See 'Further work' for possible extensions taking in account these considerations.

In Fig. 4.9 it is demonstrated that the McGM model can be used to reconstruct an input sequence. The input is a sequence of images depicting an anticlockwise rotating sine grating windmill. One of the images is shown from the input sequence, then one of the images from the sequence blurred in time and space (i.e. the windmill convolved with the standard blur kernel – the Gaussian in space and the log Gaussian in time) as calculated by the original McGM model for motion calculations (Fig. 4.9(b)). Part (c) of the figure shows the Taylor series reconstruction using the sampled blurred image and corresponding differentials. As can be seen, the lattice over which the rebuilding process is defined deals adequately with the circular shape and reproduces the fine resolution. Fig. 4.9(d) shows the scaled difference between the two images.

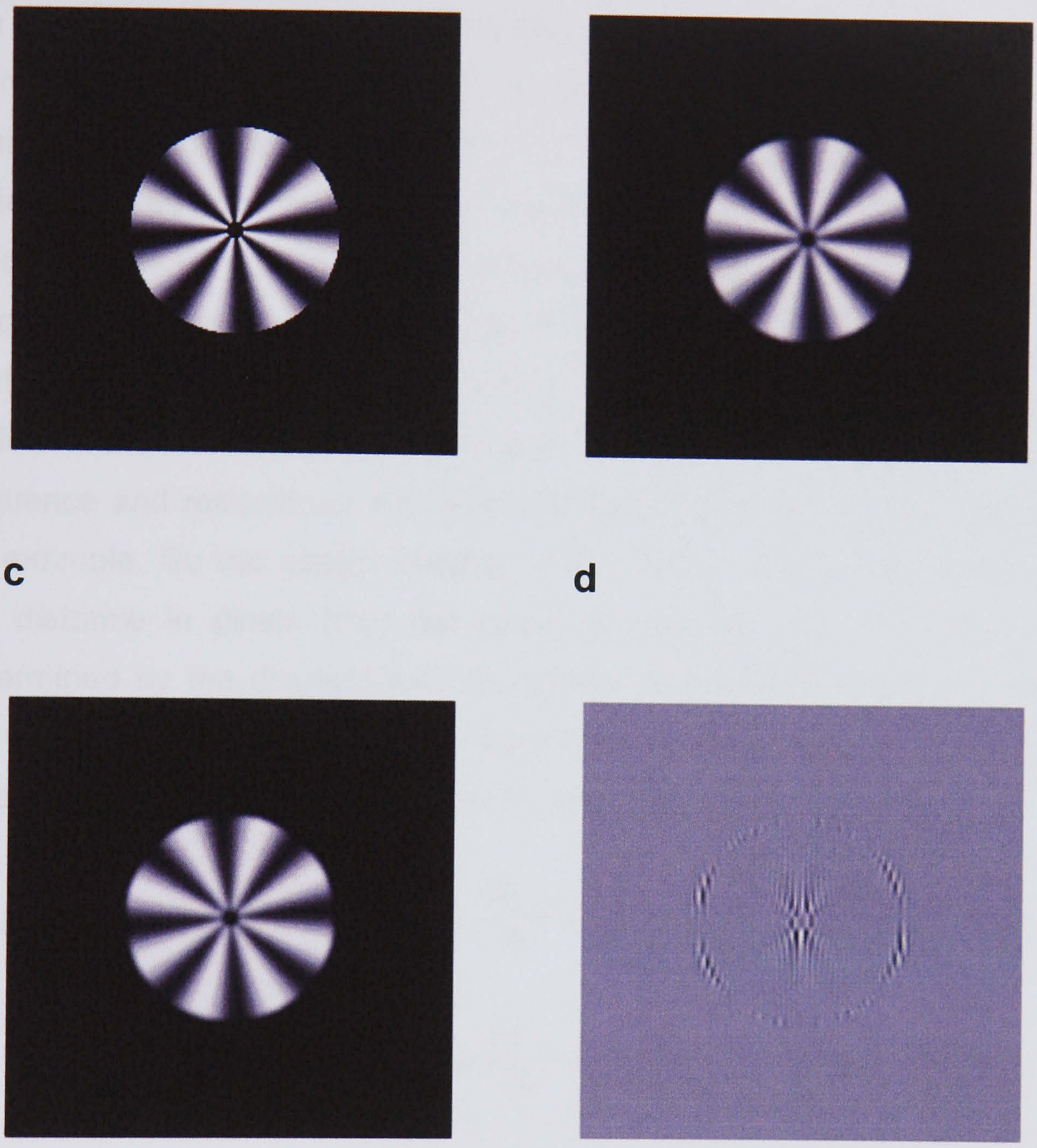


Fig. 4.9 Taylor series based reconstruction from a sequence of images using the McGM model. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor reconstruction window=3×3 pixels. (a) One of the frames taken from the input sequence of a sine grating windmill rotating anticlockwise at 1.8°/frame. (b) An example of a blurred frame calculated from the input sequence by convolving with the standard blur kernel. (c) The same blurred image represented using the Taylor series expansion calculated from derivatives of the sequence created by the motion model. (d) The scaled difference between the blurred image and the reconstruction, from (black) –1.33% difference to (white) 1.41% difference of maximum image brightness.

4.5 - Temporal and spatial representation

We have discussed how we can spatially reconstruct an image using its spatial derivatives as part of its Taylor representation. For a sequence of images with luminance changing across space and time we have also described a Taylor series in three dimensions (Eqn. 4.4), the two spatial dimensions and time, which the McGM then uses as a basis for motion calculations. Another way of reconstructing an image that is part of a sequence would be to use both its spatial and temporal derivatives. Above, we sampled the image and only took derivatives at every ninth pixel value. Similarly, we can sample the movie sequence and reconstruct the entire sequence using only every second frame for example. So the spatial weights in the Taylor equation are determined by the distance in pixels from the point of expansion and the time weight is determined by the displacement in number of frames forward or backward to the frame we wish to reconstruct from the original image (i.e. in this case the parameter only varies between 0 or 1). See Eqn. 4.13.

$$R_{i+p,j+q,k+r} = B_{i,j,k} + p\frac{\partial B_{i,j,k}}{\partial x} + q\frac{\partial B_{i,j,k}}{\partial y} + r\frac{\partial B_{i,j,k}}{\partial t} + \ldots$$

$$+\frac{1}{2!}\left[2pq\frac{\partial B_{i,j,k}}{\partial xy} + 2pr\frac{\partial B_{i,j,k}}{\partial xt} + 2qr\frac{\partial B_{i,j,k}}{\partial yt} + p^2\frac{\partial B_{i,j,k}}{\partial x^2} + q^2\frac{\partial B_{i,j,k}}{\partial y^2} + r^2\frac{\partial B_{i,j,k}}{\partial t^2}\right] + \ldots \quad (4.13)$$

$R_{i,j,k}$ = The value i pixels in x direction and j pixels in the y direction of frame k of the rebuilt output.

B = The input image sequence blurred in space-time.

When $r = 0$ in Eqn. 4.13, frame k of the blurred image sequence is reconstructed by sampling the blurred output and its spatial derivatives. When $r = 1$, the next frame ($k+1$) is reconstructed by sampling the spatial and temporal derivatives of frame k of the blurred image sequence.

When we apply this procedure to a sine grating drifting upwards at two pixels/frame, then the first frame is reconstructed just as before, as there is zero temporal offset from the support and only the spatial weights are used. In the second frame, using the temporal weight $r=1$ and the same spatial weights, each pixel value is reconstructed using only values from the derivatives of the previous frame. We get a two pixel advance, matching the speed of the grating (Fig. 4.10).

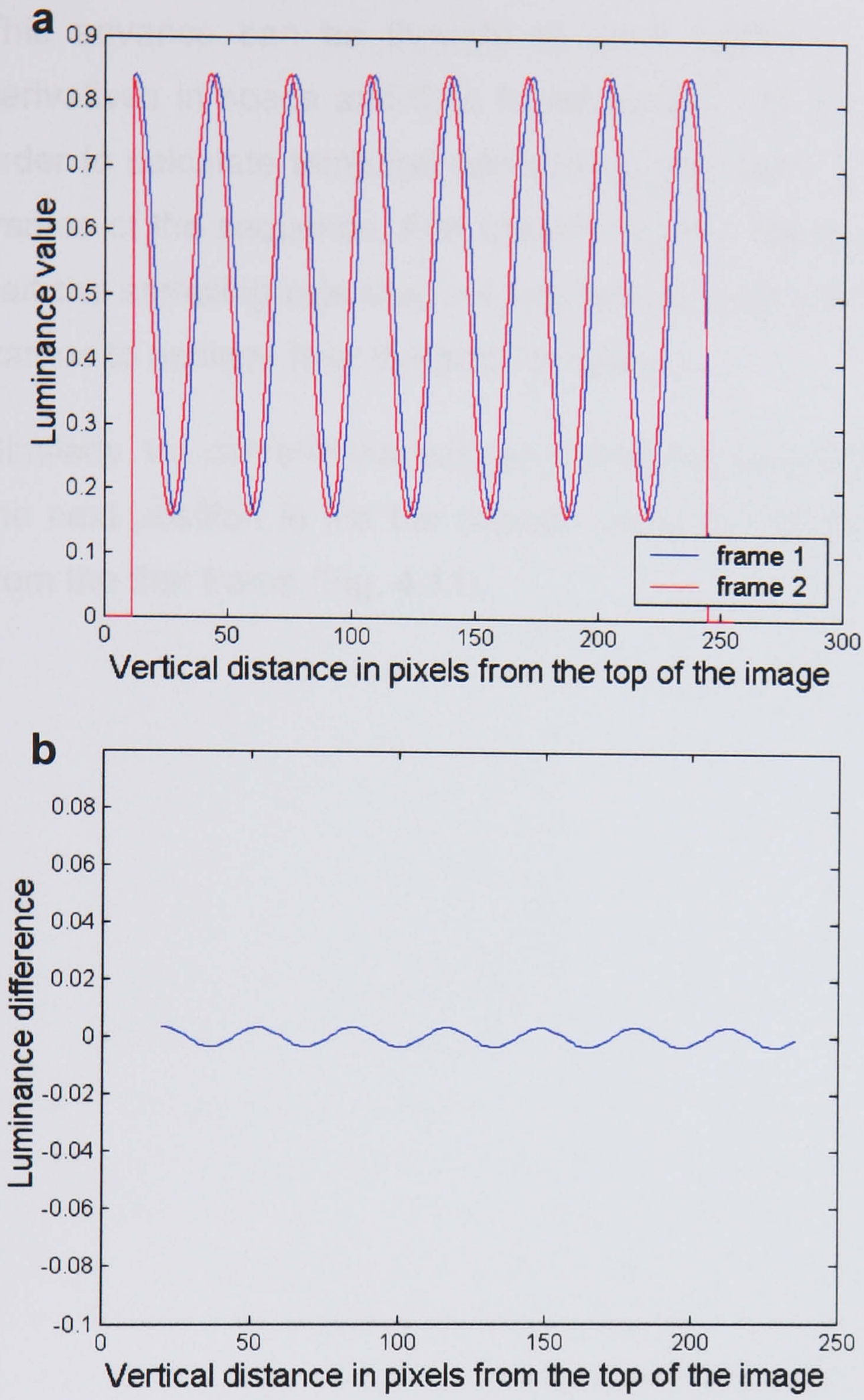
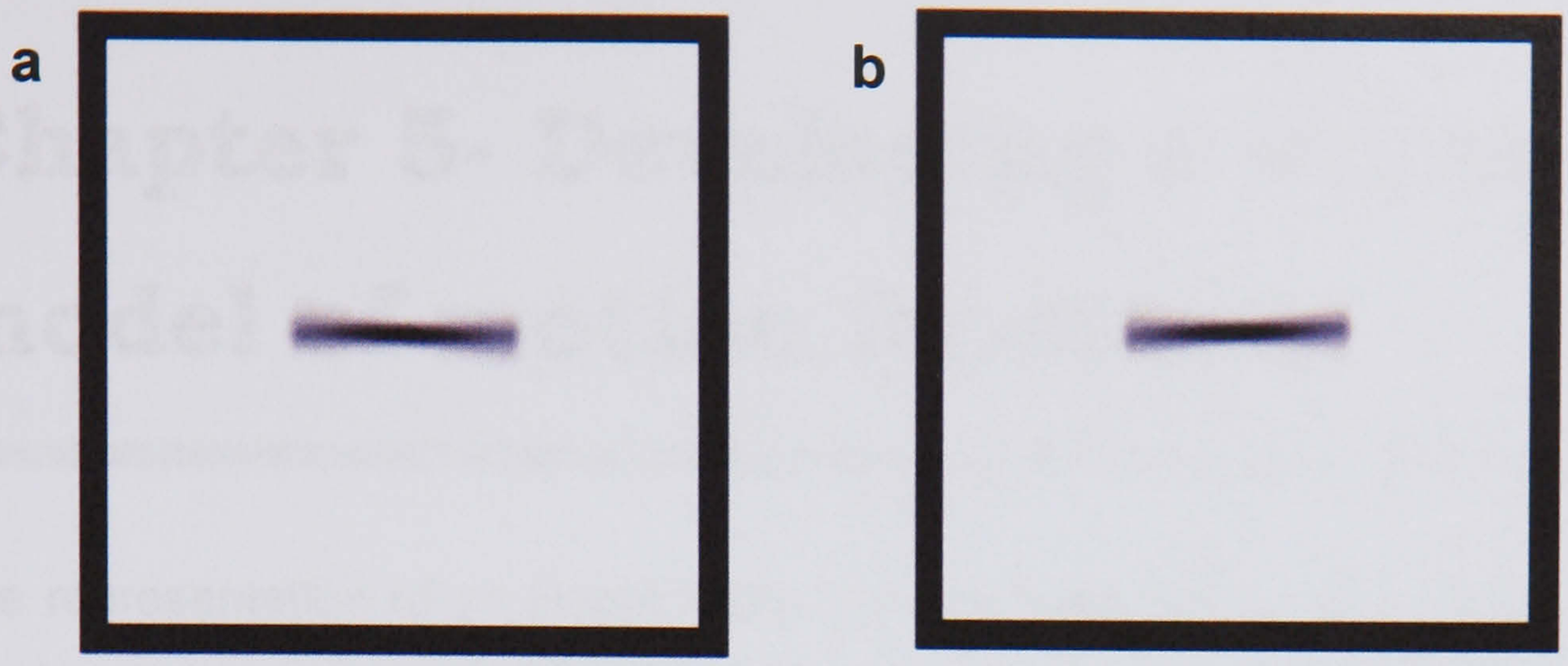


Fig. 4.10 Reconstructing a frame from a sequence of a sine grating. Taylor series based spatial and temporal reconstruction from a sequence of images using the McGM model. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, $\tau = 0.275$, temporal filter length = 23 frames. Spatial reconstruction window=3×3 pixels. (a) Vertical line of luminance values through the middle of the first and second frame of output that have been reconstructed from a sine wave drifting upwards at 2 pixels/frame using spatial and temporal derivatives from the model. (b) Difference plotted between the second frame that has been reconstructed using derivatives from the first frame and the blurred sine wave output corresponding to the second frame. Values plotted along a vertical line through the middle of the difference of the two outputs +/-0.03% max/min difference of max image brightness.

This advance can be thought of as a prediction ahead using the smooth derivatives in space and time to extrapolate the next spatial step. However, in order to calculate temporal derivatives, you need to use the previous and next frames of the sequence. Alternatively we can say that in this case we have only half the sampling rate then we had before, and we can fill in between sampled frames to achieve finer temporal resolution.

Similarly, we can see that with an anticlockwise rotating bar, we can reconstruct the next position in the bar rotation using the spatial and temporal derivatives from the first frame (Fig. 4.11).



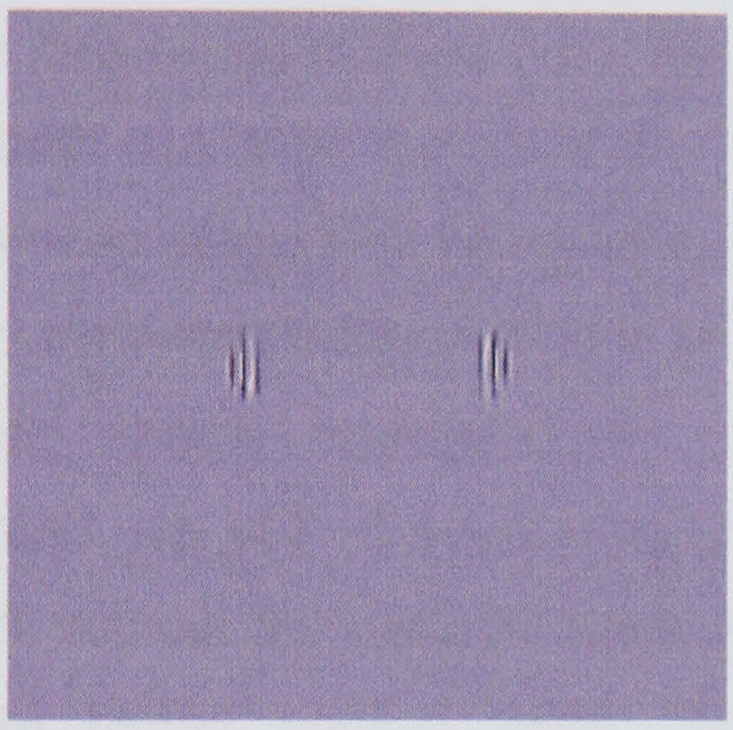


Fig. 4.11 Results from the sequence of a bar rotating anticlockwise using the McGM model. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial reconstruction window=3×3 pixels. (a) Frame from the sequence after blurring with the standard blur kernel. (b) Next frame reconstructed only using values from the frame shown in (a) and its derivatives. (c) Scaled difference between the next frame of blurred output after (a) and the reconstructed frame, +/-0.3% max/min difference of max image brightness.

In summary, we have seen how the Taylor jets can be used to represent a scene and to calculate motion as part of the McGM scheme. The work in the next chapter will concern the development of a model that can incorporate feedback from the motion calculation into the spatial representation in a way that reflects the experimental results we have described. This will be a step along the way to trying to explain the effects of visual motion on perceived spatial position.

Chapter 5- Developing a working model of motion feedback

The representation of an image using the truncated Taylor series can now be used to represent each of the images in a series of frames. At the same time motion can be calculated from the input sequence. The image representation described can then be used to incorporate motion feedback in a feasible mechanistic way. The ideas and subsequent results presented here aim to demonstrate that through a 'low-level' feedback connection between V1 where accurate spatial representation exists and MT+ areas where motion is calculated we can explain how motion induced spatial distortions could be caused, without the need for such concepts as 'grouping' and 'binding' involving different 'higher' cortical areas. The cortical representation of local luminance values is used and the way in which such a representation is formed is changed. The output of the model can be seen as V1 activity, which we are going to assume is closely linked to the final percept. There is some basis for this as it has been suggested that feedback to V1 from MT+ is a necessary condition for awareness of visual motion (McGraw, Barrett et al., 2002; Pascual-Leone & Walsh, 2001). The reconstructed image will be presented for the purpose of demonstrating perceptual effects. Again, all the stimuli presented to the model consist of sequences of 256×256 pixel images.

At first we will look at the ways of developing a model that can qualitatively reproduce the kind of motion effects we have observed in past empirical work and in the experimental work of this thesis. Once a satisfactorily realistic model has been developed, varying different parameters will test the robustness and behaviour of the model. When we have satisfied ourselves of the practical

application of such a mechanism we can investigate its limits and behaviour as part of a theoretical biological system.

5.1 - Combining motion output with spatial representation

The Taylor representation of the image depends on the weight parameters used in the summation of the derivatives of Gaussian filters' outputs. Up till now these weights have been chosen on the basis of the formula for the Taylor expansion. The great advantage of this representation is that by changing the weights in the sum of the filter outputs one can alter the representation that is formed. In this way the response in V1 is no longer locked in with the retinal image.

The labile nature of the Taylor series representation is illustrated in Fig. 5.2. The image has been reconstructed in a very similar manner to that described in the chapter before. The weights p and q, which were previously determined by the distance of a point from the centre of the Taylor expansion window in the x and y directions respectively have now been altered to produce a spatial shift.

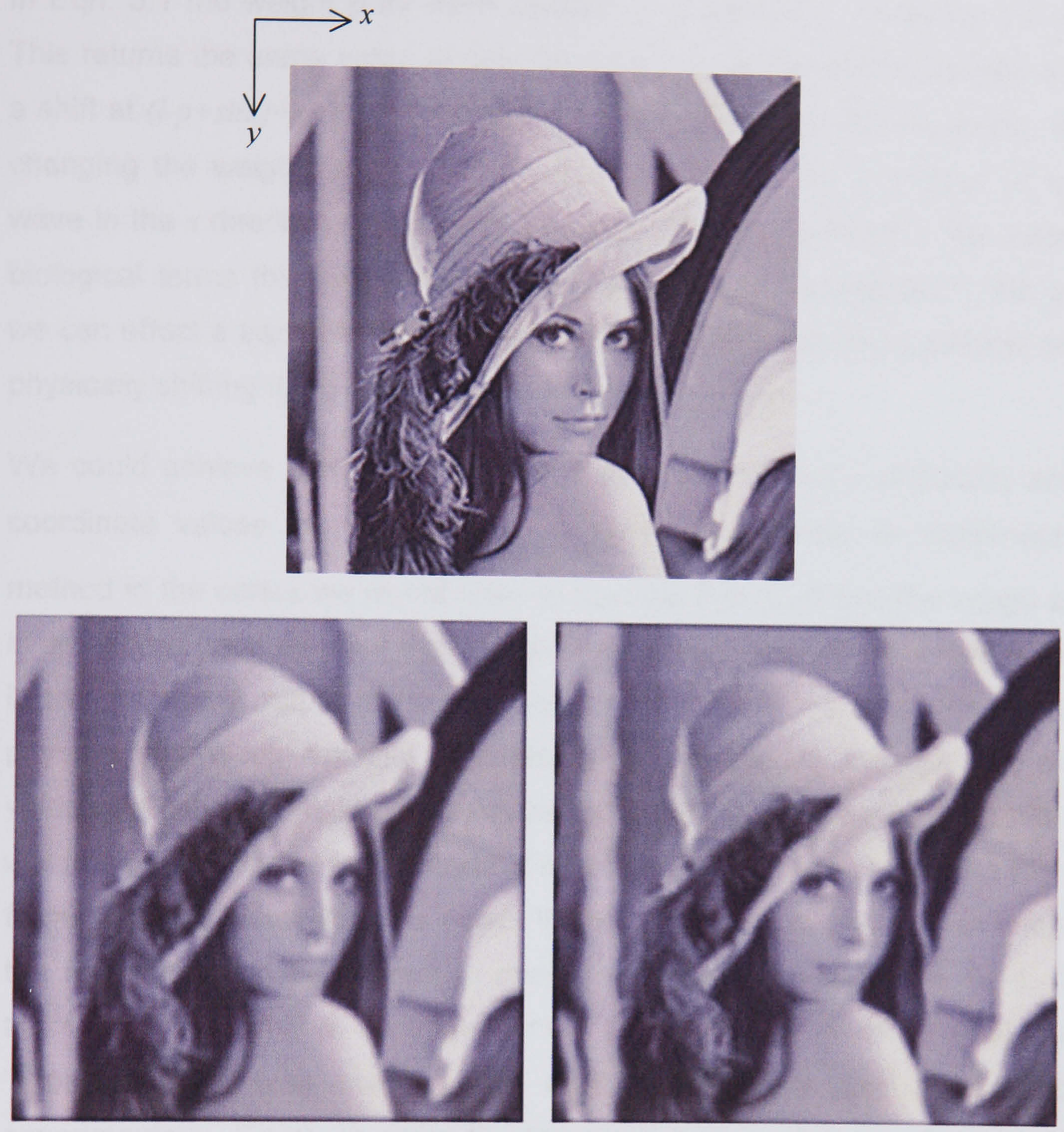


Fig. 5.2 Original image shown on top. On the left the image blurred in space with a 2D Gaussian kernel and on the right is shown the image distorted using the Taylor series reconstruction with the weights changed such that:

$$R_{i+p,j+q} = B_{i,j} + (p - \sin(j+q))\frac{\partial B_{i,j}}{\partial x} + q\frac{\partial B_{i,j}}{\partial y} + \ldots \tag{5.1}$$

$R_{i,j}$ = element at ith pixel in x direction, jth pixel in y direction of rebuilt image R

B = Gaussian blurred image

Blur support area 23×23 pixels, S.D. of Gaussian, σ = 1.5, reconstruction window 3×3 pixels.

In Eqn. 5.1 the weight p for each position (i, j) has been altered to $p\text{-}sin(j+q)$. This returns the same value at position $(i+p, j+q)$ as the value returned without a shift at $(i\text{-}p+sin(j+q), j\text{-}q)$. Crucially, the input image remains the same, but by changing the weights in the Taylor representation by the amplitude of a sine wave in the x direction we observe a corresponding translation in the output. In biological terms this means that by changing the representation in the cortex we can effect a translation that would give rise to the equivalent percept as say physically shifting the piece of paper we were observing.

We could achieve the same effect in the output by simply physically shifting coordinate values i.e. $M_{i,j} \rightarrow M_{i+\sin(j),j}$. However, in order to implement this method in the cortex we would need to assume first of all that the image exists in an explicit point-by-point expression of image brightness rather than implicitly in the combined output of filters. Secondly, it is hard to envisage how such a physical shift would actually be implemented. How does one cell take on the value of a cell two neighbours away for example? It would seem that this step would actually require some physical re-wiring, which is not feasible in the time frame in which visual illusions occur. Without an actual change of the image on the retina it would be difficult to implement such a cortical shift. Also, if the translation were only to be implemented locally such a re-wiring would cause 'gaps' in the visual scene – no such visual effect has been reported experimentally. Finally in order to implement such a shift two neural 'sheets' would be needed to represent both the input and the shifted position as it is calculated.

With the implementation that uses the Taylor series representation these problems do not arise. The same 'hard coding' remains of the visual scene, but the weightings are altered on each of the cells that fires. The shift is achieved 'in place' with no need to have access to nearby values. In order to utilise the Taylor series approximation, the image does however have to meet the criterion of 'smoothness' (i.e. no abrupt changes in luminance) – the blurred image conforms to this requisite.

A local shift can be achieved by changing the weights locally and no 'gaps' are created in this process. A useful analogy is that of sub-pixel motion. If we wanted to create several frames in a sequence of a moving grating we could move the grating one pixel upwards per frame, by re-designating each pixel value as the one below it after each frame presentation. If we want to produce motion that is slower than one pixel per frame we can instead calculate the values of the pixels for a sine grating that is shifted in phase by less than a pixel. Instead of translating the value of one pixel to another, by changing the brightness of each pixel by a smaller amount than the translation step the effect of sub-pixel motion is created. See Fig.5.3.

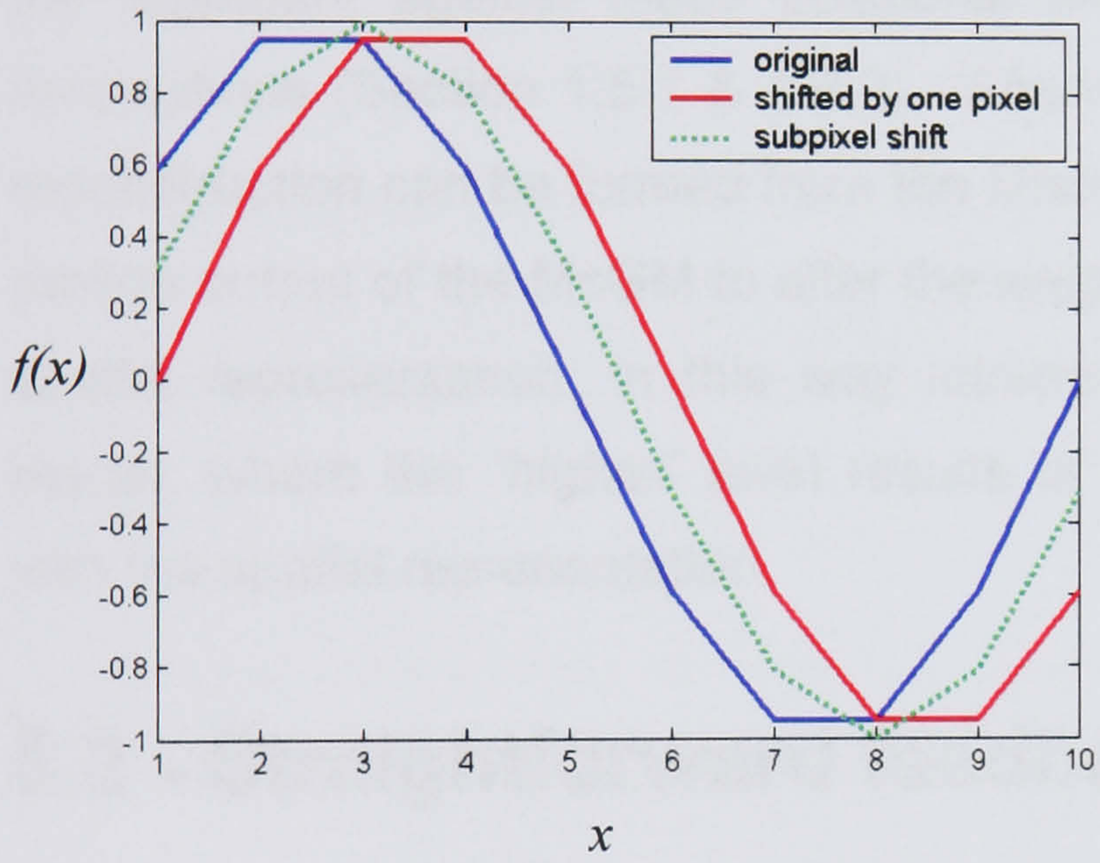


Fig. 5.3 For $x = 1{:}10$ the original sine wave is given by $S(x) = \sin(x*2*pi/10)$, the shifted sine wave is given by $Sshift(x) = S(x\text{-}1)$, wrapped round at the beginning. The sub-pixel shifted sine wave is given by $S_{sub}(x) = \sin(x*2*pi/10 + 2*pi/20)$ a sine wave phase shifted by a 20^{th} of the cycle, i.e. half a pixel.

In a similar way, in the Taylor jet method of reconstruction the values at each pixel are slightly increased or decreased to produce the effect of a translation. A limit to the size of shift that can be implemented in this way is that a Taylor approximation only yields accurate results within a given neighbourhood from the point of approximation. This means that the size of the translation implemented in this way can only be small, relative to the size of the image. This limitation however fits in well with the size of these shift effects as

observed in psychophysical experiments, as the shifts are only around 2-10 arc minutes compared to stimulus sizes of 1°-9° (De Valois & De Valois, 1991; Whitney & Cavanagh, 2000).

Within the context of the Taylor jet based spatial representation, the question is in what way these weights become influenced in the presence of motion. The weights could be altered by horizontal inhibitory/excitatory connections known to exist in V1 and used as the bases of previous dynamic models of V1 activity (Hirsch & Gilbert, 1991; Li, 2001). This would involve the weights on the filter outputs being linked to the outputs of other filters. However, it is difficult to see how motion would influence this activity per se and we have already discussed the argument against these positional shifts being mediated by V1 lateral connections (Section 1.5.1 & 1.5.2). Alternatively, as we have shown that the reconstruction can be formed from the filters of the McGM, so we could use the motion output of the McGM to alter the weights of the components that form the spatial representation, in this way introducing a feedback connection in the model, where the "higher" level results of motion processing are recombined with the spatial representation.

5.2 - Straightforward feedback (Version 1)

The motion model can be adapted to reconstruct blurred versions of the input images at the same time as calculating velocity. In this section we will concentrate on ways of altering the spatial representation, using only the two dimensional spatial Taylor expansion, not the temporal aspect, to see to what extent translations can be reproduced purely using a purely spatial manipulation. For each image blurred in space and time there is also a corresponding velocity output. Next, the way in which the spatial representation is constructed is changed so that the velocity values are incorporated into the Taylor weights. Where above we introduced a shift parameter of $\sin(j+q)$, I now simply introduce the corresponding velocity value at pixel (i, j). This model

assumes that for each V1 receptive field there is a corresponding calculated motion value, which then boosts or inhibits the receptive field response accordingly.

So, the Taylor representation now takes on the form:

$$R_{i+p,j+q} = B_{i,j} + (p - \xi u)\frac{\partial B_{i,j}}{\partial x} + (q - \xi v)\frac{\partial B_{i,j}}{\partial y} + \ldots \quad (5.2)$$

$R_{i,j}$ = pixel i,j of the rebuilt image R, i in the x direction, j in the y direction

B = blurred image

ξ = multiplier for the amount of feedback

u = velocity component in the x direction

v = velocity component in the y direction
The components of the velocity in each direction are given by

$$u = \cos(A_{i+p,j+q}) \cdot V_{i+p,j+q},\ v = \sin(A_{i+p,j+q}) \cdot V_{i+p,j+q}$$

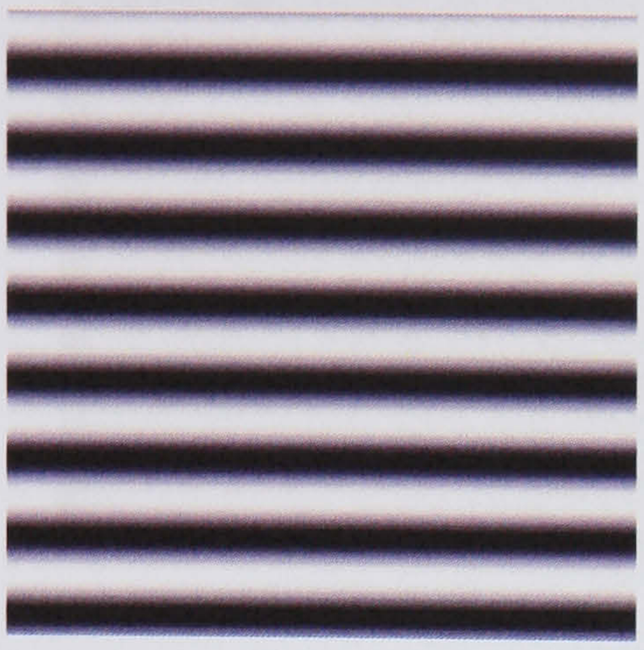
A = angular output from the McGM model for the given blurred image B giving the pixelwise velocity direction

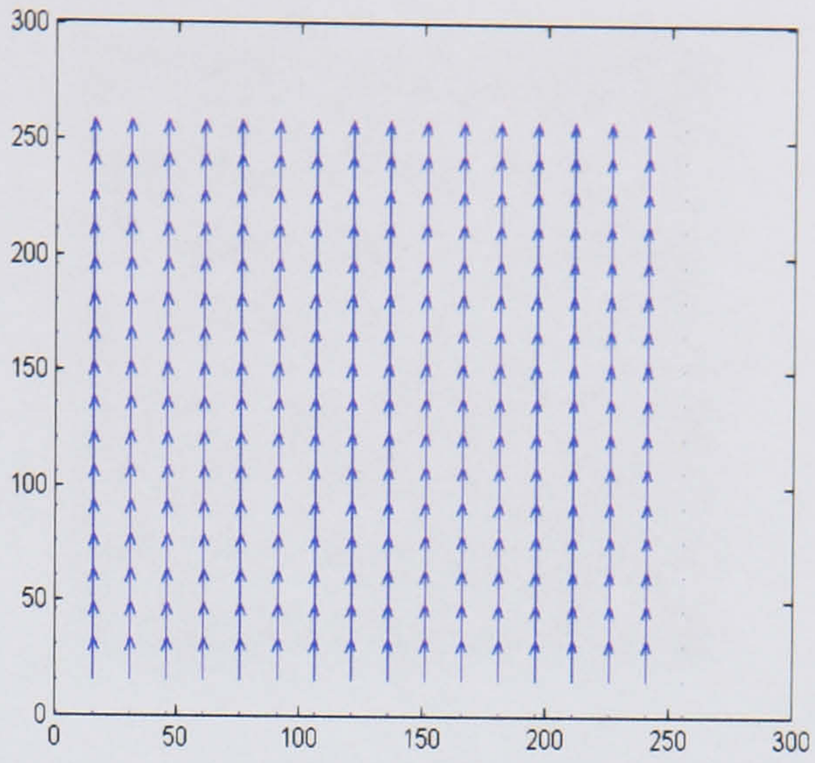
V = speed magnitude output from the McGM model for the given blurred image B

The proportion of velocity that feeds back in, ξ, will be examined later, but for now, it will assume the value of 1, i.e. exactly all of the motion feeds back into the representation. Once we have found a model that qualitatively reflects the experimental results we have discussed we can examine the effect of varying ξ. It is now trivial to achieve a shift in the direction of image motion. Note, in order to incorporate the velocity values in the weights the Taylor representation needs to be calculated after motion processing. This order of processing reflects the suggested pattern of feedback in the cortex. In Fig. 5.4. the induced shift on a blurred output image from a moving sine grating is demonstrated.

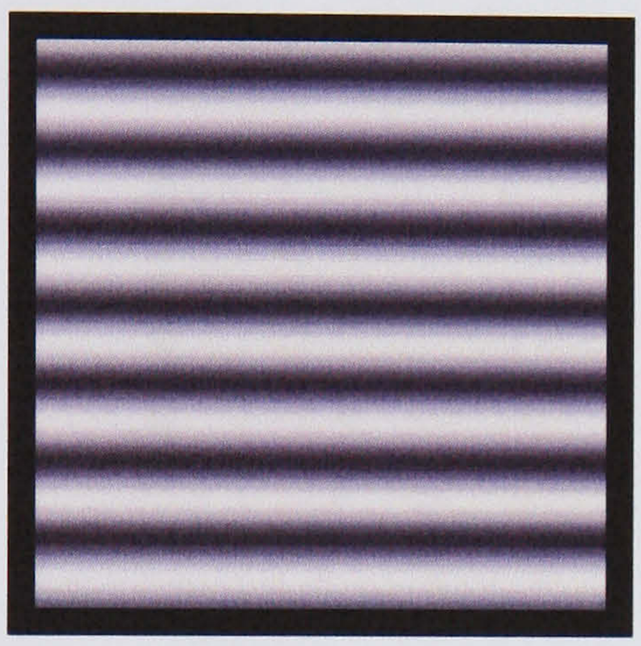
Shown is a frame from the original sequence, the blur in space and time of the sequence, the corresponding velocity output from the model and the rebuilt blurred image without motion feedback and with motion feedback, shifted in the direction of motion by the pixel level re-calculation of the Taylor series incorporating the motion output. We see that without the motion feedback the sine pattern is reconstructed fairly accurately, with the feedback the sine pattern is shifted 2 pixels in the direction of motion (phase difference found by fitting sine curves to the outputs).

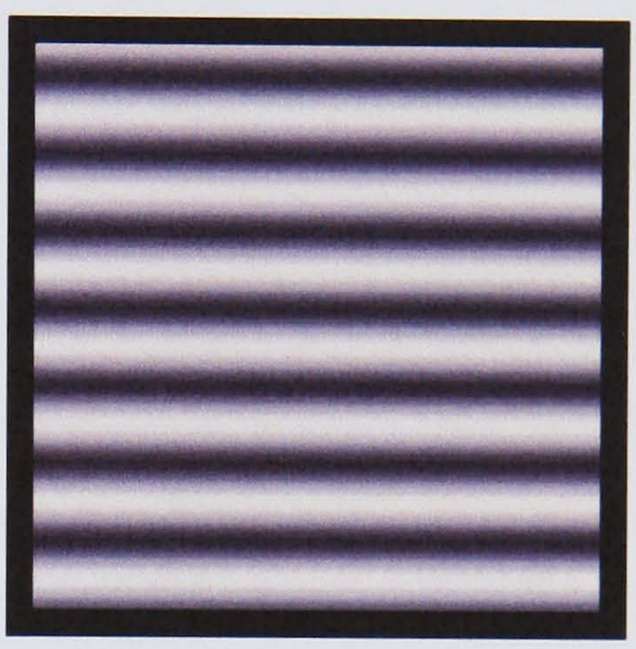
a



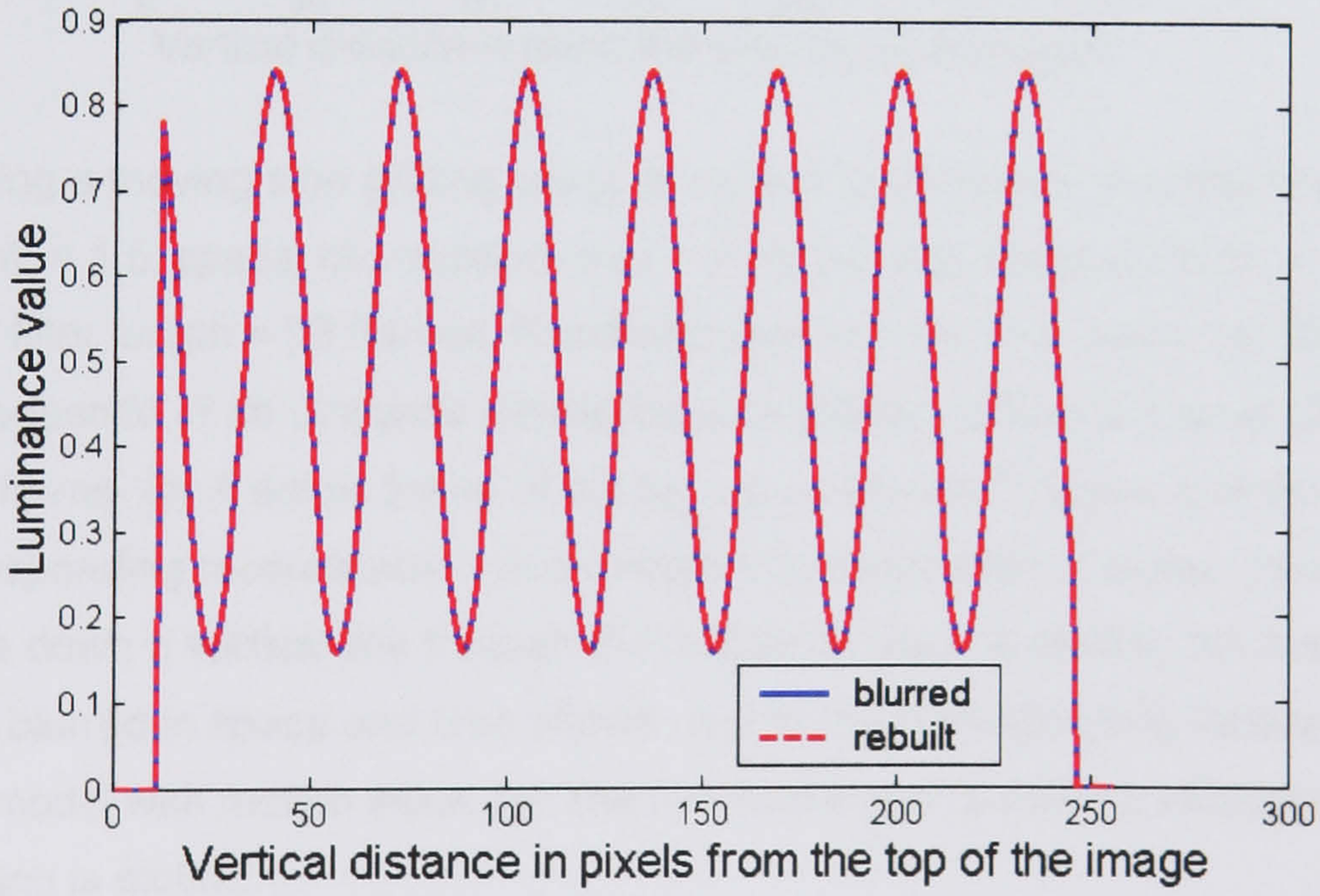

300
250
200
150
100
50
0
0
50
100
150
200
250
300

b





c


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Luminance value
0
50
100
150
200
250
300
blurred
rebuilt
Vertical distance in pixels from the top of the image

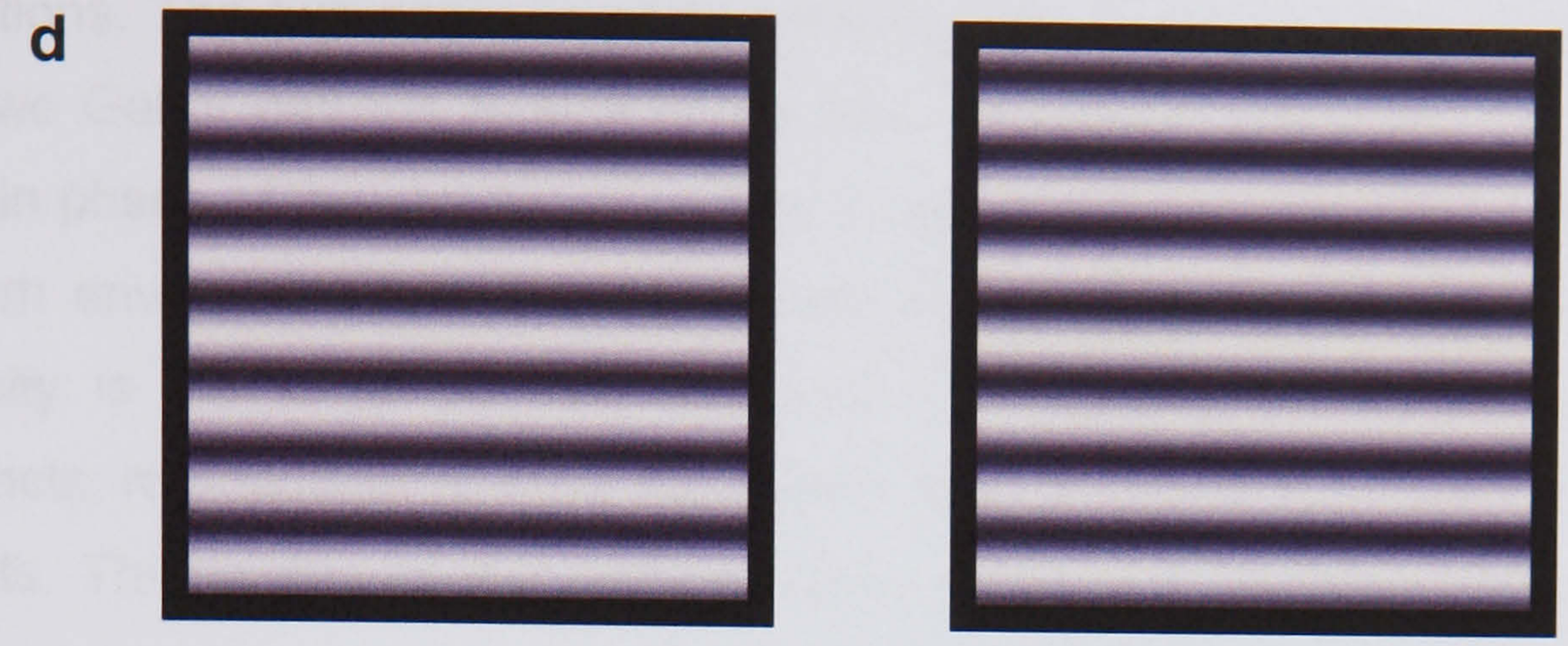


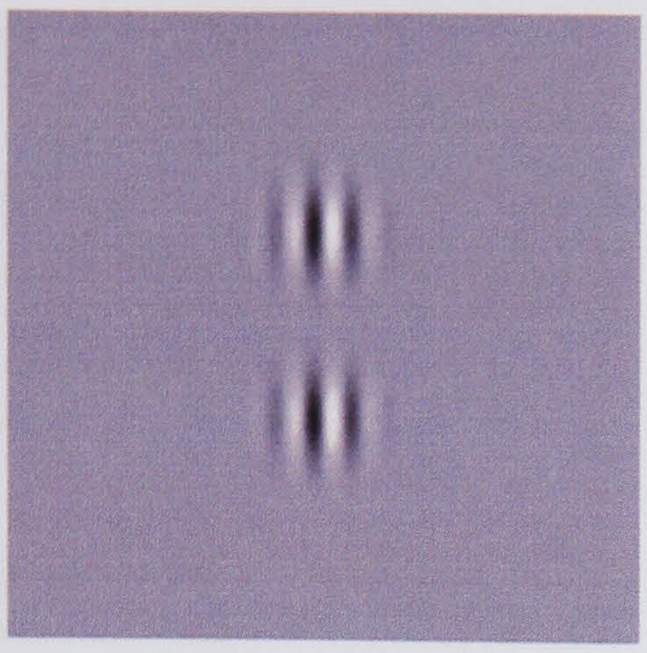
Fig. 5.4 Rebuilding a moving sine grating using its motion as feedback. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Reconstruction window 3×3 pixels. (a) Single frame from the input sequence of an upwardly translating sine grating (2 pixels/ frame) shown next to a motion output frame. (b) A single frame of the sequence blurred in space and time shown next to the corresponding reconstruction output from the model without motion input. (c) The luminance profile down a vertical line through the middle for each is plotted. (d) A single frame of the sequence blurred in space and time shown next to the corresponding reconstructed output from the model with motion input. (e) The luminance profile down a vertical line through the middle for each is plotted.

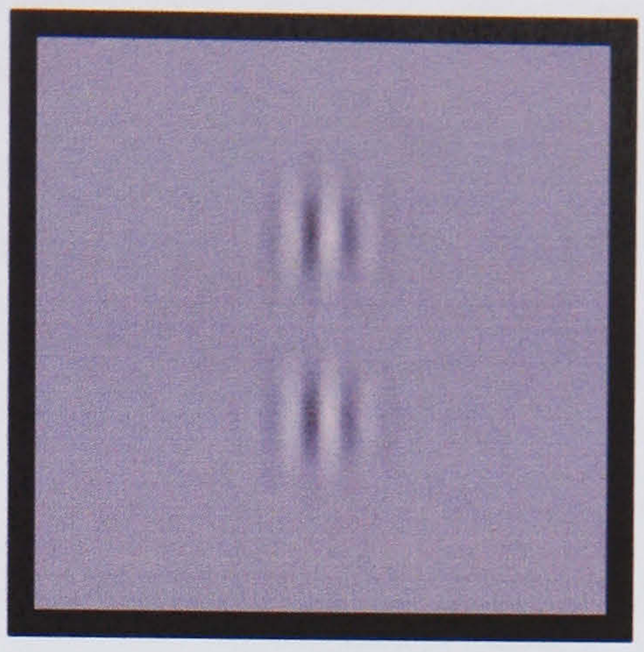
This trivial shift can also be used to reproduce the De Valois and De Valois (1991) effect, in which a perceptual shift is observed between the static envelopes of two patches containing Gabor patterns drifting in opposite

directions. The two opposite motion fields produce velocity values that cause the two Gabor patches to shift apart. See Fig. 5.5. In this case it is not just a shift in phase of the sine wave pattern as we saw above, but also a shift of the pattern envelope. These velocity values should be homogeneous as the drift velocity is the same all over the pattern. The model does produce some artefacts, reproducing some of the pattern and causing some inconsistent edge effects. This is due to the static window. There is a reduction or increase in contrast that is correlated with the change in position i.e. it is not a truly rigid motion. When the horizontal line is plotted through the centre of each of the two blurred Gabor patches reconstructions (with motion feedback and without), we can see that the latter output is misaligned in the direction consistent with the motion. If Gabor curves are fitted to each of these profiles, whereas there is no difference in their midpoints for the straightforward reconstruction, for the reconstruction with feedback there is a difference of about 4 pixels, in the direction of motion, which is consistent with the drifting velocities of the patches that determine the shift parameter.

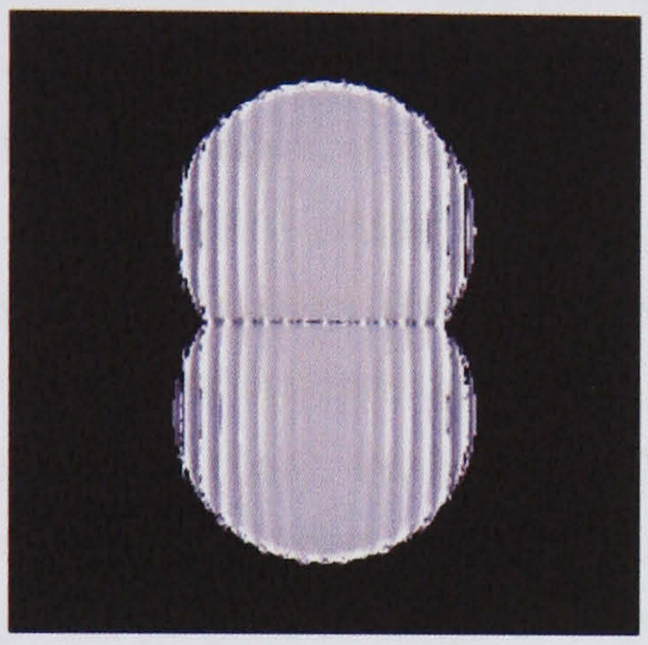
The drifting Gabor patterns have not changed as input stimulus and at any given time their envelopes are aligned on the retina. What we are changing is the way we represent the input by incorporating the motion information. The output corresponds to the observer's percept of shifted patches. By fitting Gabor functions to the output we find that the two envelopes have shifted by 4.61 pixels w.r.t. each other.

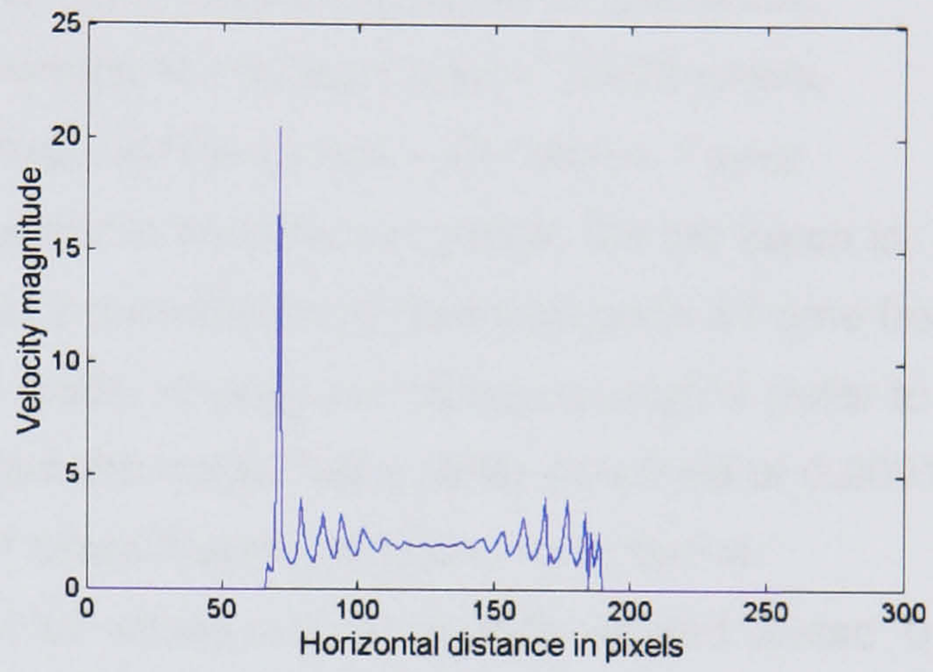
a

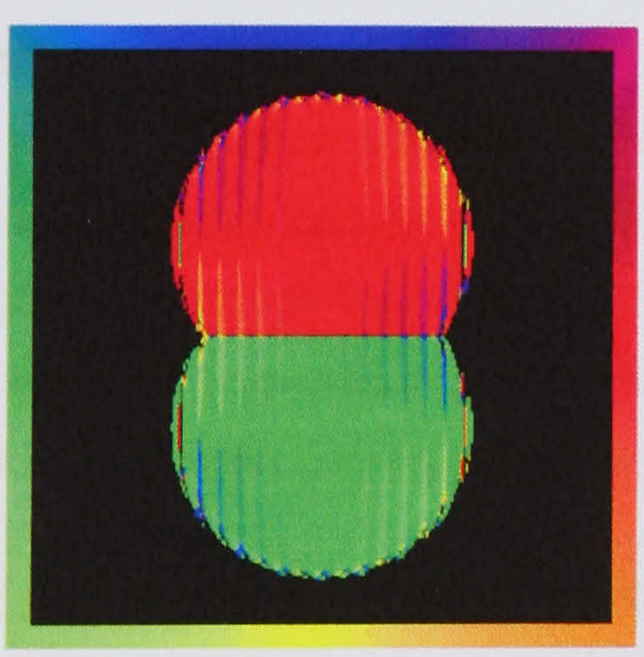




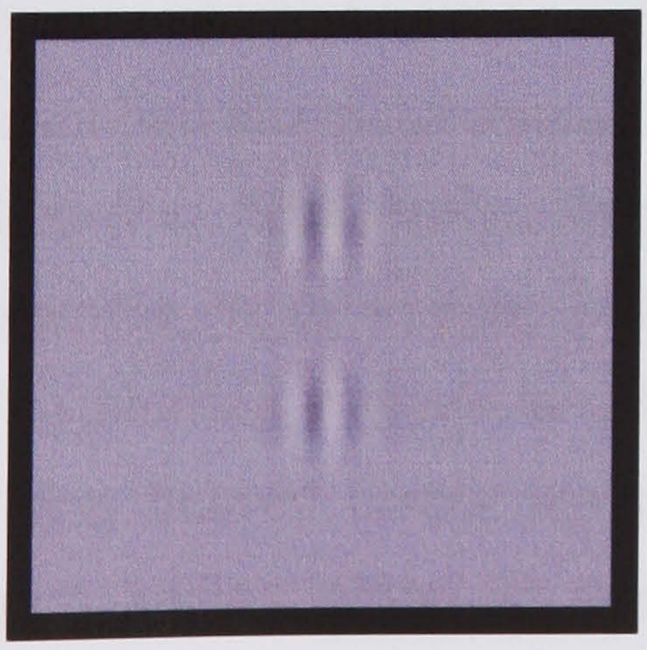
b

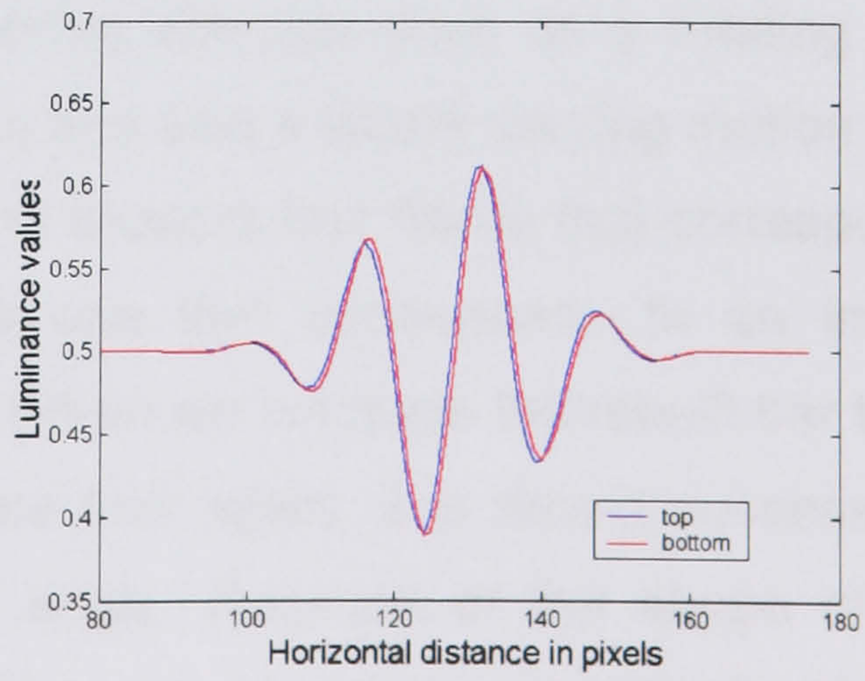



25
20
15
10
5
0
0
50
100
150
200
250
300
Velocity magnitude
Horizontal distance in pixels



c




0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
80
100
120
140
160
180
Luminance values
Horizontal distance in pixels
top
bottom

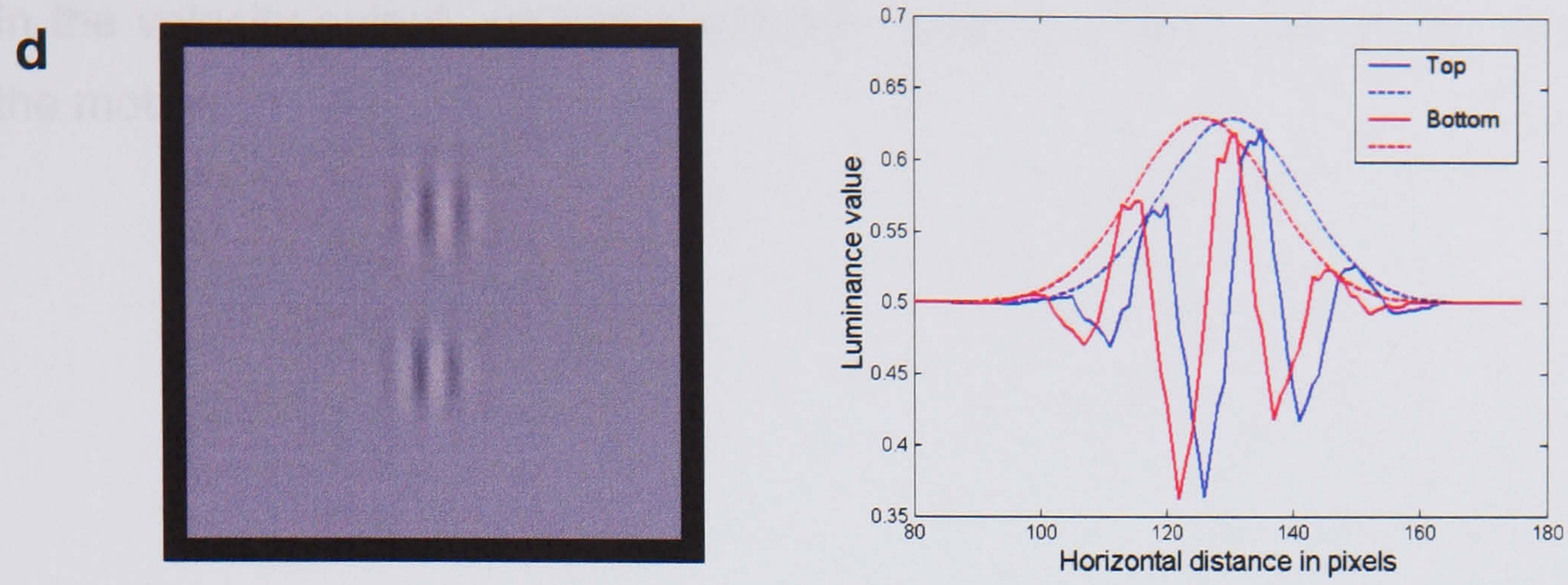


Fig. 5.5 Illustration of the De Valois and De Valois shift effect as predicted by the model. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) A single input frame from the sequence, the top patch is drifting right and the bottom patch is drifting left at 2 pixels/frame shown alongside a frame from the sequence blurred in space and time. (b) The model velocity and direction outputs (refer to colour wheel for direction). Velocity magnitude from the model has a lower threshold of 0.0001 pixels/frame. Only directions for motion over 0.01 pixels/frame are shown in all further illustrations. The magnitude image is thresholded for values over 3 pixels/frame and scaled 0 (black) - 3 (white). Note that the output is not completely homogeneous as might be expected from the motion input. The velocity magnitude is plotted for the horizontal line through the top patch, the values are around 2 pixels/frame except for an anomaly at the edge. (c) One of the images from the sequence blurred in space and time. The blue line illustrates luminance along the horizontal line through the middle of the top patch, the red line through the bottom one. (A luminance value of 1 corresponds to white, 0 to black) (d) The rebuilt image corresponding to the blurred input image. Luminance values plotted as above.

The sine grating and drifting Gabor patches have uniform motion fields across time and smoothly varying luminance levels across space. Let us now consider a spatially discrete but continuously moving stimulus such as a rotating bar. This stimulus has sharp luminance edges and also a locally varying motion field through time (Fig. 5.6). Again, there is no exact output frame that corresponds to the example input image, but only one that corresponds to an image produced by blurring in space and time. When we compare the rebuilt bar to its corresponding blurred output we can see that again, it is rebuilt successfully and appears shifted along by a small angle. Because of the shape of the motion field the rotating bar is shifted along in a way that preserves its shape.

In the velocity output, we can again see some anomalies around the edges of the motion.

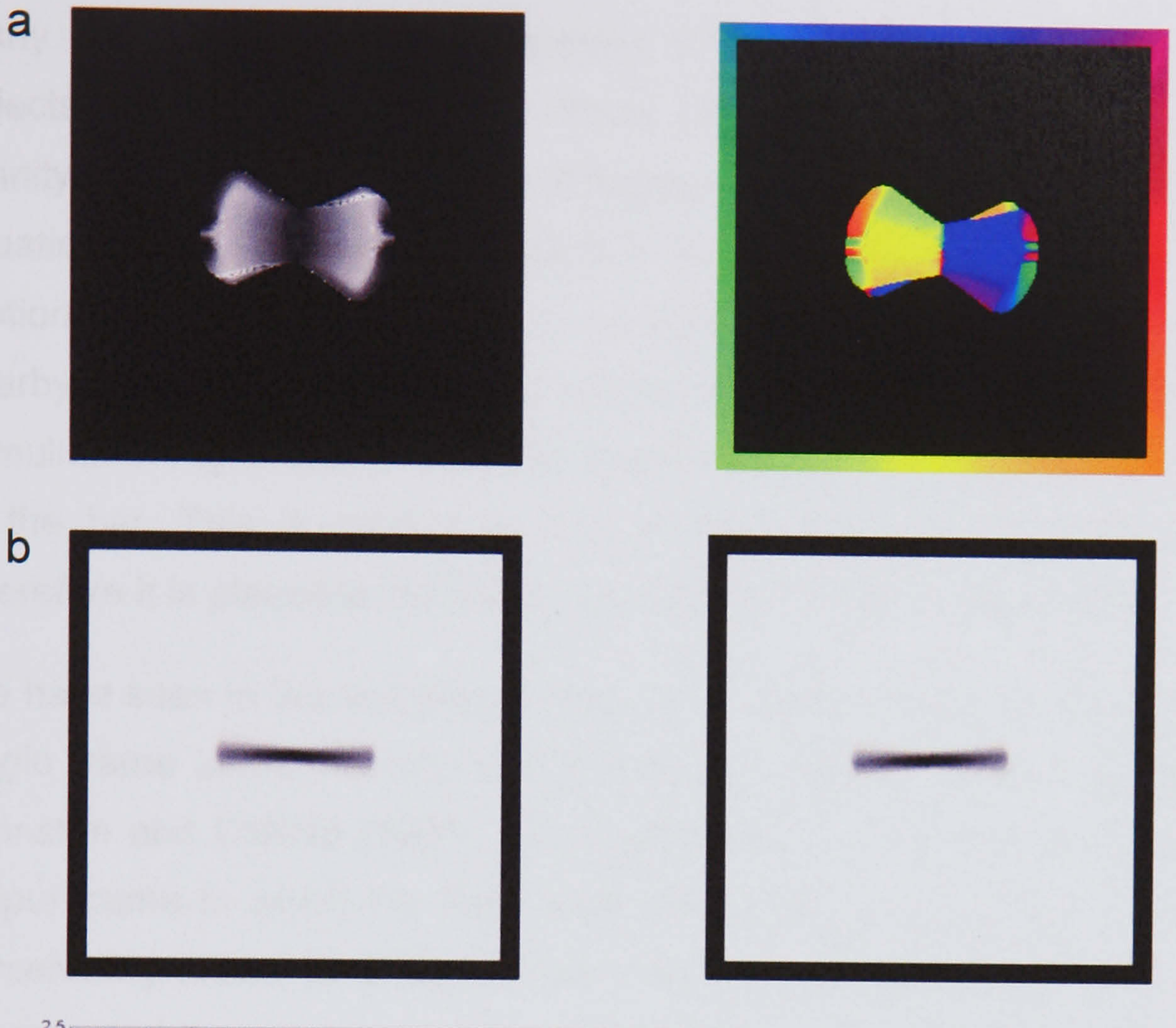


Fig. 5.6 Rotating bar sequence input into the first version of the feedback model, bar rotating anticlockwise at 3°/frame. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Single frame of velocity output from model, showing magnitude and direction. Magnitude is thresholded for values over 3 pixels/frame and scaled 0(black) – 3(white). (b) The rotating bar blurred in space and time and the corresponding reconstructed image using the velocity values as weights in the Taylor representation. This is shifted ahead. (c) A plot of a horizontal line through the middle of the image sampled when the bar is horizontal, showing input motion versus the motion calculated by the model.

Many of the experiments described above involve short presentations of objects, which then disappear. Small flashed bars, when presented in the vicinity of motion, appear to be dragged along with it. This is different to the situation above where a shift occurs in the position of the stimulus creating the motion field. In order for a flash to be shifted not by its own motion, but motion nearby, motion signals need to extend beyond the boundary of the moving stimulus. In Fig. 5.6(c) we can see that the motion does extend beyond the end of the bar. This is caused by the spatial integration in the motion model. Therefore it is plausible that local motion could affect nearby objects.

We have seen in the last chapter that the model response to a single flash in a single frame takes the form of the temporal impulse function as described in Johnston and Clifford (1995), and implemented in the McGM. Location in the output frame in which the flash response peaks can be taken to indicate the perceived position of a flash. This may not accurately reflect the biological process, but the process of decision making is beyond the scope of this thesis.

We have already observed that a briefly presented flash has a highly irregular signature in space-time. It has no orientation in space-time and very high temporal derivatives as it appears and disappears discretely with no continuous build up. This makes this straightforward pixel-wise motion calculation approach unsuitable for use in the position coding of a single flash. Moreover, in this version of the model the influence of motion cannot extend suitably far from the boundaries of the moving object. These two problems are illustrated below with the grating stimulus used by Whitney and Cavanagh (2000). Under experimental conditions the flashes are observed to be misaligned from each other in the direction of motion closest to them. Fig. 5.7 shows the input frame in which the flashes appear. Although the single frame with the flashes present is symmetrical, the two gratings are drifting in opposite directions as can be observed in the velocity field shown. The blurred output frame in which the flash response peaks is shown as well as the corresponding motion field. We can see that, although the motion of the gratings does extend beyond the

boundaries of the gratings, it does not overlap the positions of the flashes even though they are nearby. Hence, the velocity around the flashes is the same as without the nearby motion. This causes problems in the reconstruction. First of all, we know that the Taylor reconstruction only works with small values of the weights p, q as it only holds true within a given neighbourhood of the point at which we expand. Introducing these high velocity values produces inaccurate values outside the luminance range of the input image into the reconstruction. Secondly, the motion values are no longer consistent with the shape of the object causing the motion. We can see that the flash is spread out over an area much larger than the spatial blur. The shift in the flashes will follow the pattern of velocity we saw in the last chapter, so that as the velocity fluctuates they will either get spread out as shown or not affected at all when there is no velocity present around them (which is the case in frame 10, when the flash representation peaks). Note that of course the gratings are shifted in opposite directions.

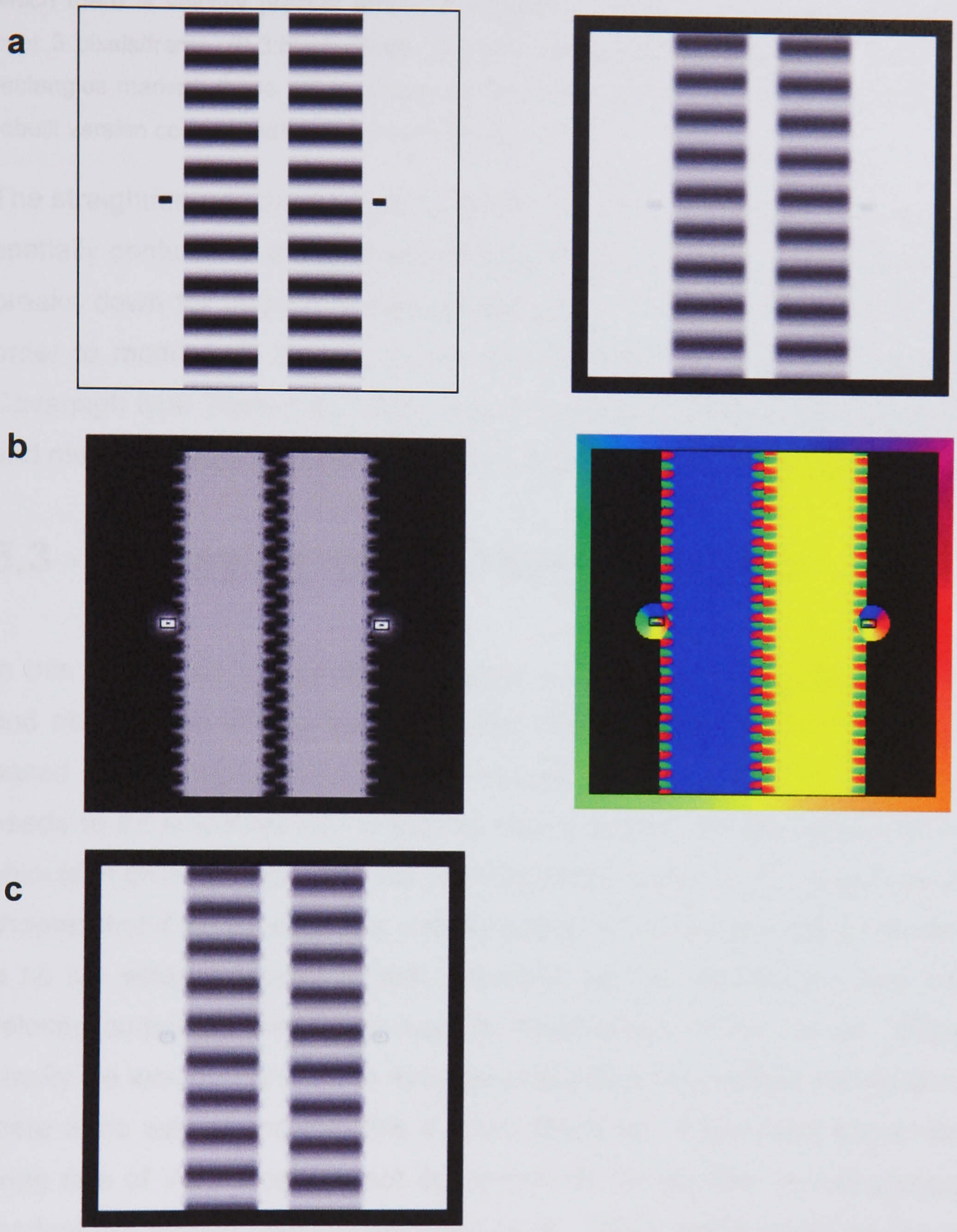


Fig. 5.7 Output from the first version of the model for an input sequence with two flashes presented horizontally in line in one frame, either side of two sine gratings translating at 2 pixels/frame in opposite directions (left grating upwards, right grating downwards). Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) The single image from the input sequence in which the flashes occur and the corresponding image blurred in space-time. (b) The velocity and directions for frame 11 in

which there is velocity present around the flashes. The magnitude is thresholded for values over 3 pixels/frame. (0-3:black-white). The pixels occupied by the flashes lie inside the black rectangles marked; these will be shown in all future velocity outputs involving flashes. (c) The rebuilt version corresponding to the velocity frame.

The straightforward pixel-by-pixel feedback of motion information works for both spatially continuous and spatially discrete motion, but the spatial representation breaks down for sudden abrupt appearance and disappearance of objects. In order to model both the De Valois and De Valois shift and the Whitney and Cavanagh type 'flash-drag' shift using the same low-level spatial representation and motion feedback model, a modification of this original idea is necessary.

5.3 - Averaging over motion calculation

In this section we are going to consider a way of modifying the motion model and observe the effects that a modified motion output has on the final Taylor based spatial representation of a sequence of images. The motion output needs to be smoother and to extend further beyond the boundaries of moving objects in order to reproduce the experimental effects. We have seen in the last chapter that if we average the velocity output of the current motion model there is no net velocity around a flash. However, we will not average over the final velocity output as this would lead to inaccuracies in the motion calculation. Ideally we would only want to average over pixels that contain movement – but there is no way of knowing this *a priori*. Moreover, it has been shown that the firing rate of V5/MT cells is not dependent on the number of dots present in a random dot motion display (Snowden et al., 1992), which would be the case if motion were summed spatially.

However, one can implement pooling at an earlier point in the motion calculations. In the McGM motion calculations we have seen that both the motion magnitude and direction are calculated from ratios of derivatives (Section 4.2). The idea would be to perform a pooling of values that contribute to this motion calculation. By applying averaging the aim is to cancel out the

local effects of a brief flash presentation, which we have seen lead to a breakdown of the spatial representations. Importantly this idea ties in with the proposed explanation of the observed time lag in the peak perceived misalignment as described in the empirical chapter. It was suggested that this was caused by a larger motion cell contributing to the output of a smaller V1 cell. By spatial averaging over the motion measures, motion further away from a pixel will contribute to its reconstructed luminance value. Biologically this is based on the fact that MT+ receptive fields have been found to be much larger than V1 cells (Mikami et al., 1986). There are various averaging steps that would smooth the motion output. As we compare the results of averaging at different stages of the model I will highlight the level by showing the stage in the box diagram from the previous chapter – the overview of the model – that we are considering.

5.3.1 Pooling over 'speed' and 'inverse speed' (Version 2 & 3)

Use the terms from matrix M to calculate speed $\hat{\boldsymbol{s}} = (\hat{s}_{\parallel}, \hat{s}_{\perp})$ and inverse speed $\check{\boldsymbol{s}} = (\check{s}_{\parallel}, \check{s}_{\perp})$

In order to retrieve the velocity value at each point in the image the model initially computes speed ($\hat{s}_{\parallel}$), orthogonal speed ($\hat{s}_{\perp}$), inverse speed ($\check{s}_{\parallel}$) and orthogonal inverse speed ($\check{s}_{\perp}$). These measures are used for calculating both the final velocity magnitude and direction. Each of these is calculated for every pixel in the image and hence forms an image size matrix corresponding to the input image. The first averaging strategy involves averaging over each of these four inputs into the final velocity calculations. These image size matrices will be filtered with a uniform filter of some given size $n \times n$ with elements $1/n$, so that the output is normalised. First of all we need to check that this modification does not affect the accuracy of the velocity calculation. The modified version was tested with a spatial pooling window over motion of 11×11, 31×31, 51×51 pixels on a moving sine grating input of size 256×256 pixels on each frame and drifting two pixels per frame upwards. All three averaging areas return a speed

of 2 pixel/frame (accurate to 4 s.f.) in a 90° (upwards) direction for all of the image area.

We can observe the effect of this averaging strategy on the De Valois and De Valois stimulus sequence. In Fig. 5.8 we can observe that when using a 11×11 averaging area, as compared to the original motion output, the velocity results are more accurate. Both the speed and direction values are more uniform with none of the slight anomalies along the luminance contours. However, in the original motion output around the edge of the Gabor patches there were just a few high velocity values. In the smoothed version this band of high values becomes wider around the edge of the Gabor patches. This is not a result of thresholding, as removing thresholds in the motion model still results in these high velocity values around the edges of motion. It seems that high edge values are due to the sharp decrease in the denominator values of the speed calculation caused by averaging over a large area where most of the speed estimates are zero. If we compare this with the larger averaging area of 31×31 pixels (Fig. 5.9), we can see that this area of high velocities becomes even larger. However, as in this case the high velocity area does not overlap with the Gabor patches in the reconstruction, we can still observe a clear shift with a realistic reconstruction (with a few pixels miscalculated due to large velocity values).

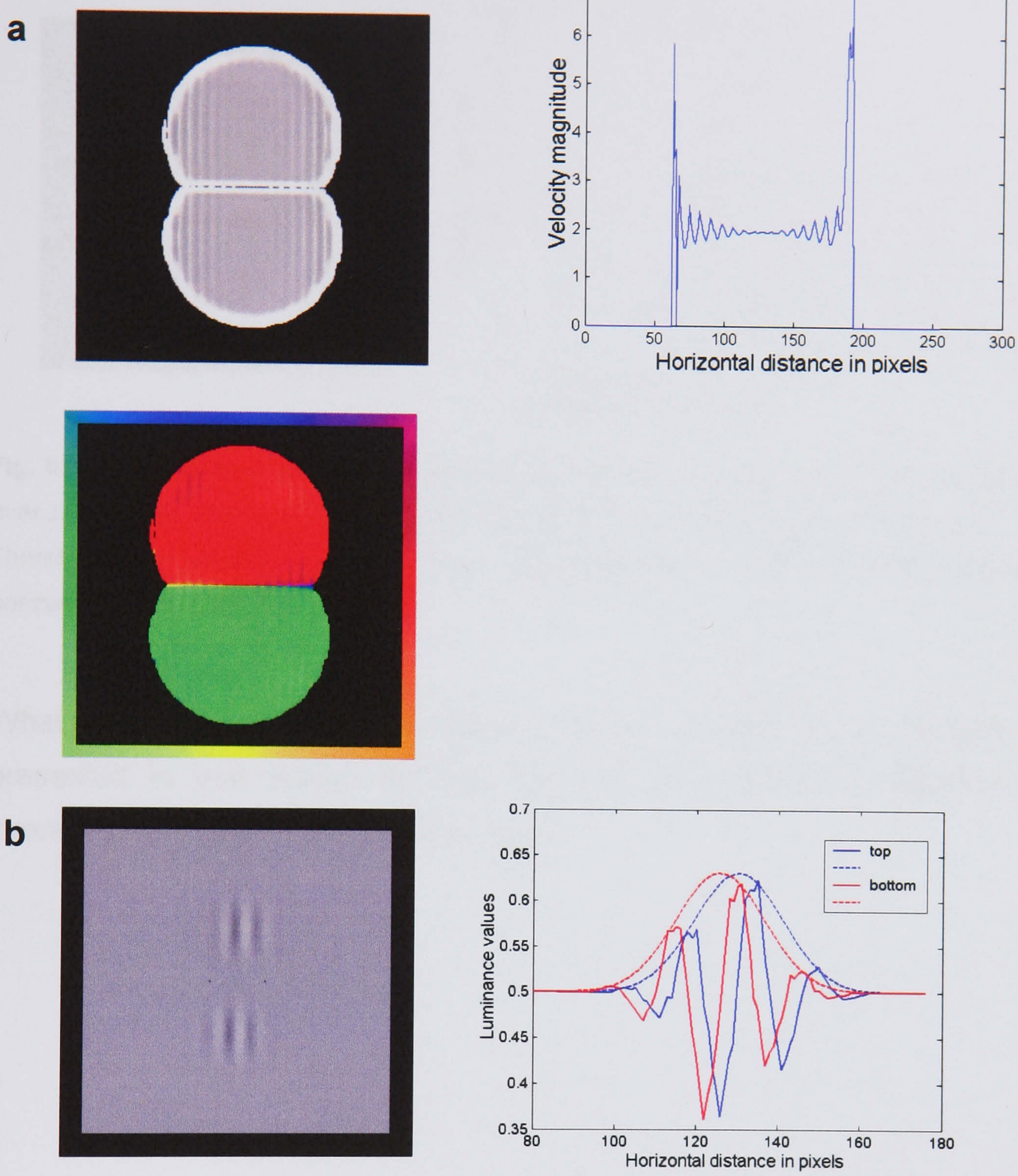


Fig. 5.8 Drifting Gabor patterns as input for the 2nd model with uniform averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 11×11 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels (a) Output velocity magnitude (thresholded above 3 pixels/frame - white, 0 - black) and direction. The correct magnitude of 2 pixels/frame is returned with a more smooth output except for some high values around the edges of each patch of motion. The velocity magnitude is plotted along the horizontal line through the top patch. (b) The Gabor patches are trivially shifted away from each other when using the motion values as parameters in the Taylor reconstruction. By fitting a Gabor to each set of pixel values, the difference is 4.5 pixels.

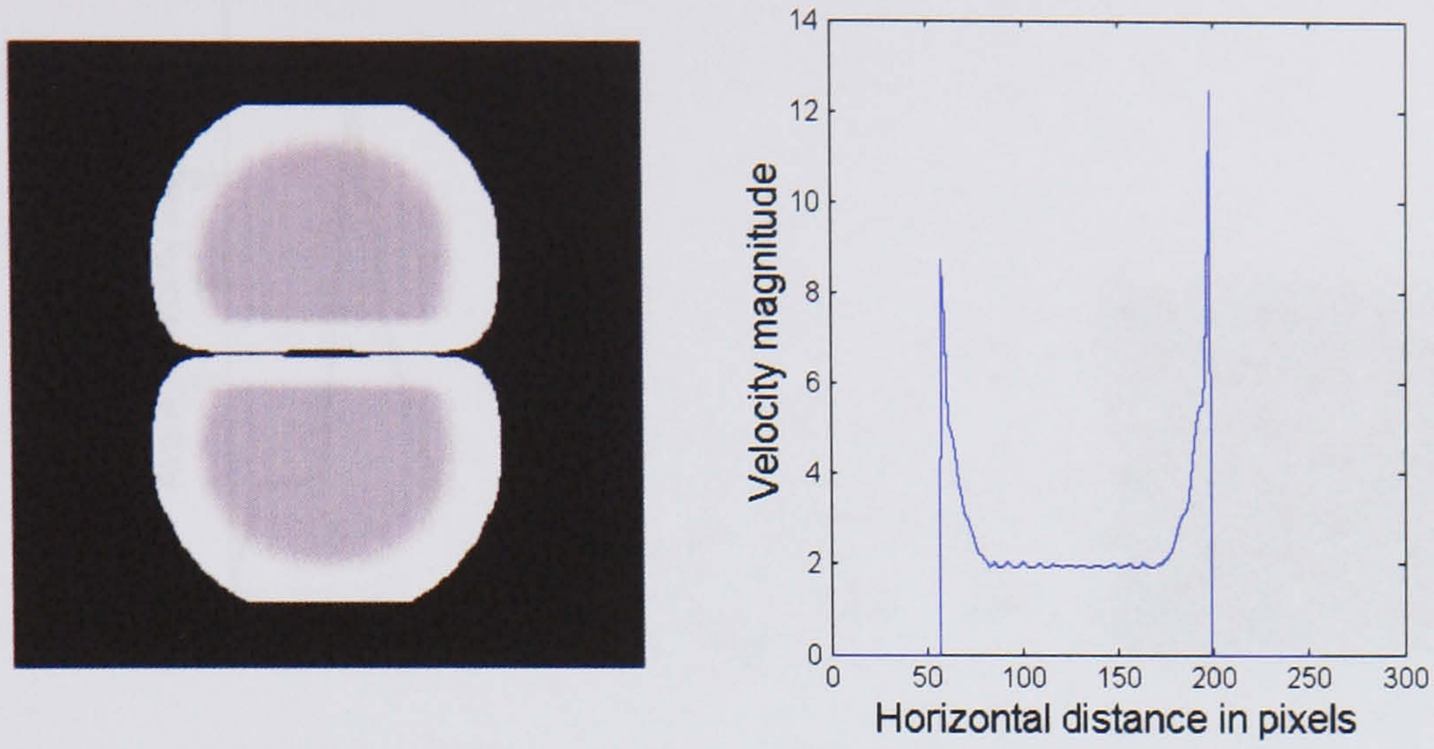


Fig. 5.9 Velocity output for drifting Gabor patches into the 2nd model with uniform averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 31×31 pixel areas. Other parameters as in Fig. 5.8. Thresholded for values over 3. (0-3: black – white). The velocity magnitude is plotted along the horizontal line through the top patch.

What response does this averaging model provide for a discrete flash presented in one frame? In Fig. 5.10 we see results for different sized averaging windows on the velocity values.

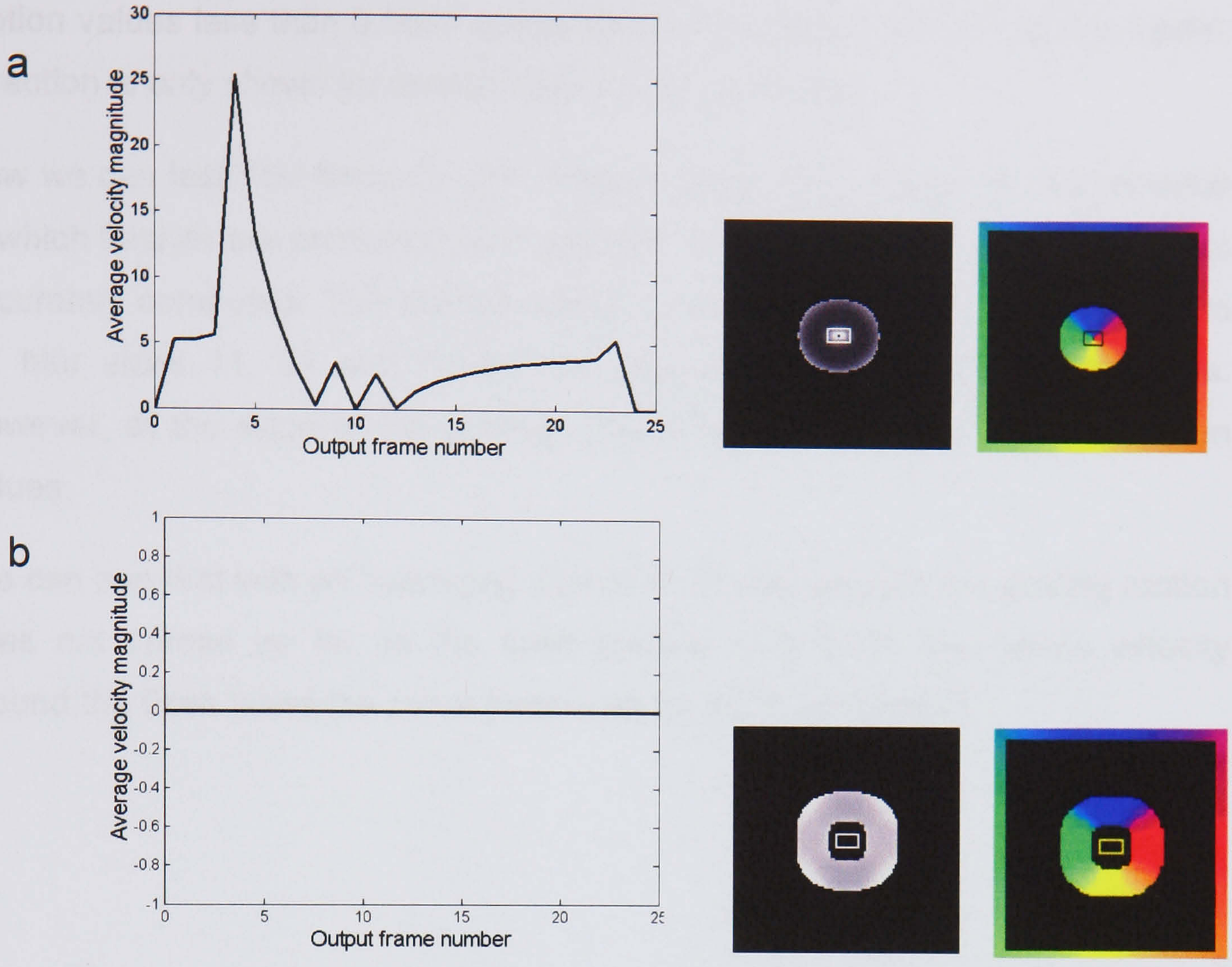


Fig. 5.10 Velocity outputs for 2nd feedback model with averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\check{s}_{\perp}$. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. The average magnitude over the area of the flash is plotted. Also shown is the spread of the velocity magnitude and direction for frame 11 in which velocity is present. Thresholded for values over 5. (0-5: black – white). (a) Velocity blur = 11×11 pixels. (b) Velocity blur = 31×31 pixels.

We can see that for the 11×11 pixel blur the average velocity magnitude remains the same although the pattern of the velocity is slightly changed. However, with the 31×31 pixel blur, there is no velocity present at the flash position. The surround velocity varies in magnitude in the same pattern as before, i.e. in frame 10, when the flash representation peaks there is no velocity at all present. With a 51×51 pixel size blur, the velocity associated with the flash completely disappears, with no velocity present in any of the frames. Note that the strange shape of the velocity field is due to the thresholding, whereby

motion values less than 0.0001 pixels/frame are returned as zero by the model. Direction is only shown for motion over 0.01 pixels/frame.

Now we can test how these flashes interact with nearby motion for the stimulus in which flashes are presented near gratings. First, we check the speed is again accurately computed. The correct speed, 2 pixels/frame is returned as before for blur sizes 11, 31 and 51 for the area containing the moving gratings. However, at the edge of the gratings this averaging introduces higher motion values.

We can see that with an averaging size of 11 at this distance the grating motion does not spread as far as the flash position (Fig 5.11) and hence velocity around the flash takes the same pattern as for the flash by itself.

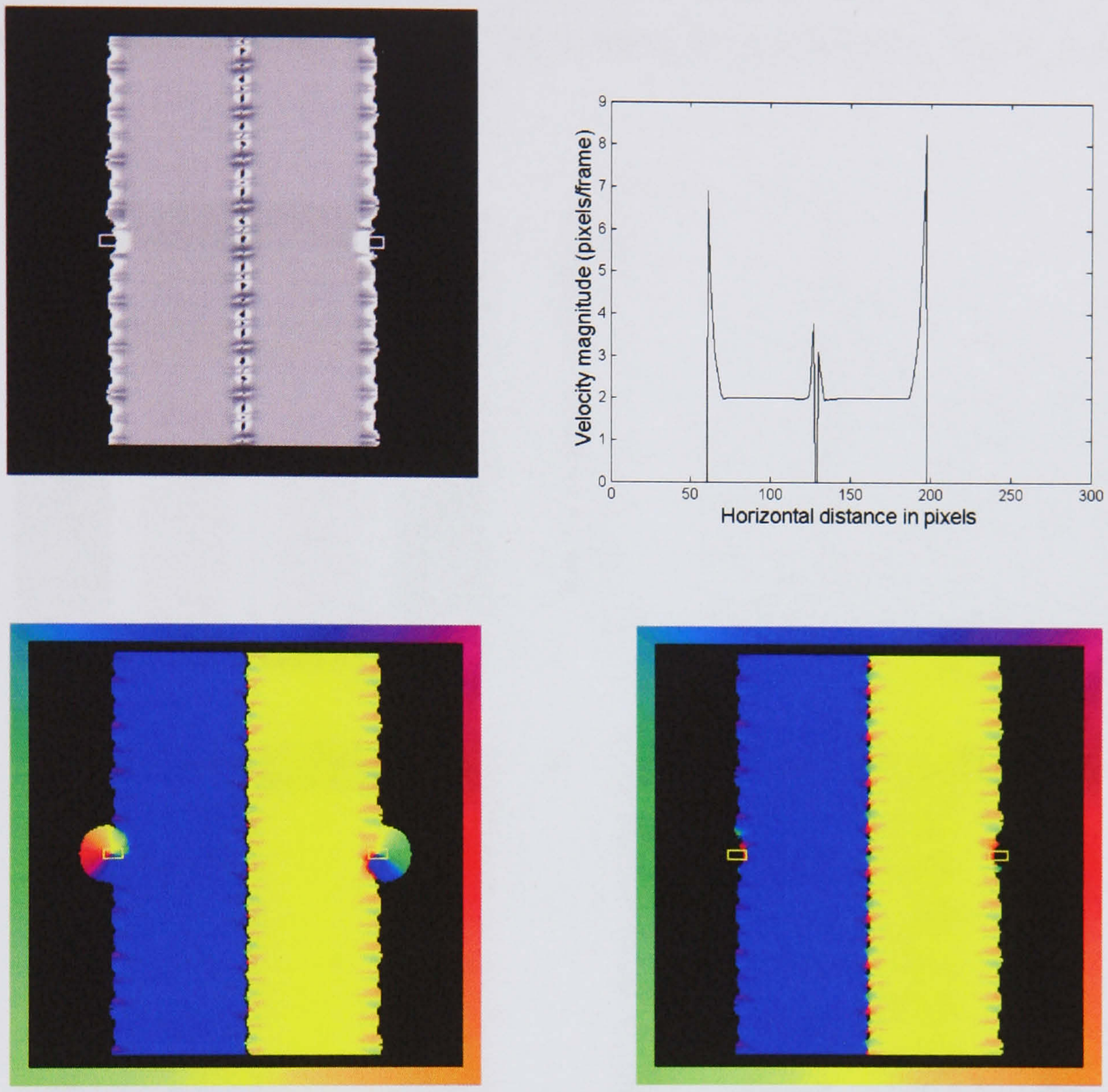


Fig. 5.11 Inputting two oppositely moving gratings (left upwards, right downwards), with flashes either side into the 2nd model with uniform averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 11×11 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude from a frame with no velocity present around the flash. Thresholded for values over 3. (0-3: black – white). Also plotted are the velocity values along a horizontal line through the middle of the gratings. (b) Motion direction from a frame with motion present around the flash and from without motion present around the flash.

However, if the larger averaging window size of 31 is used, we see that the consistent grating motion is spread out further, over the area occupied by the flashes. However, in this implementation, the motion at the edges of the

gratings takes large values. The result is poor reconstruction, causing the flash values to become distorted, and generating an unrealistic image (See Fig. 5.12).

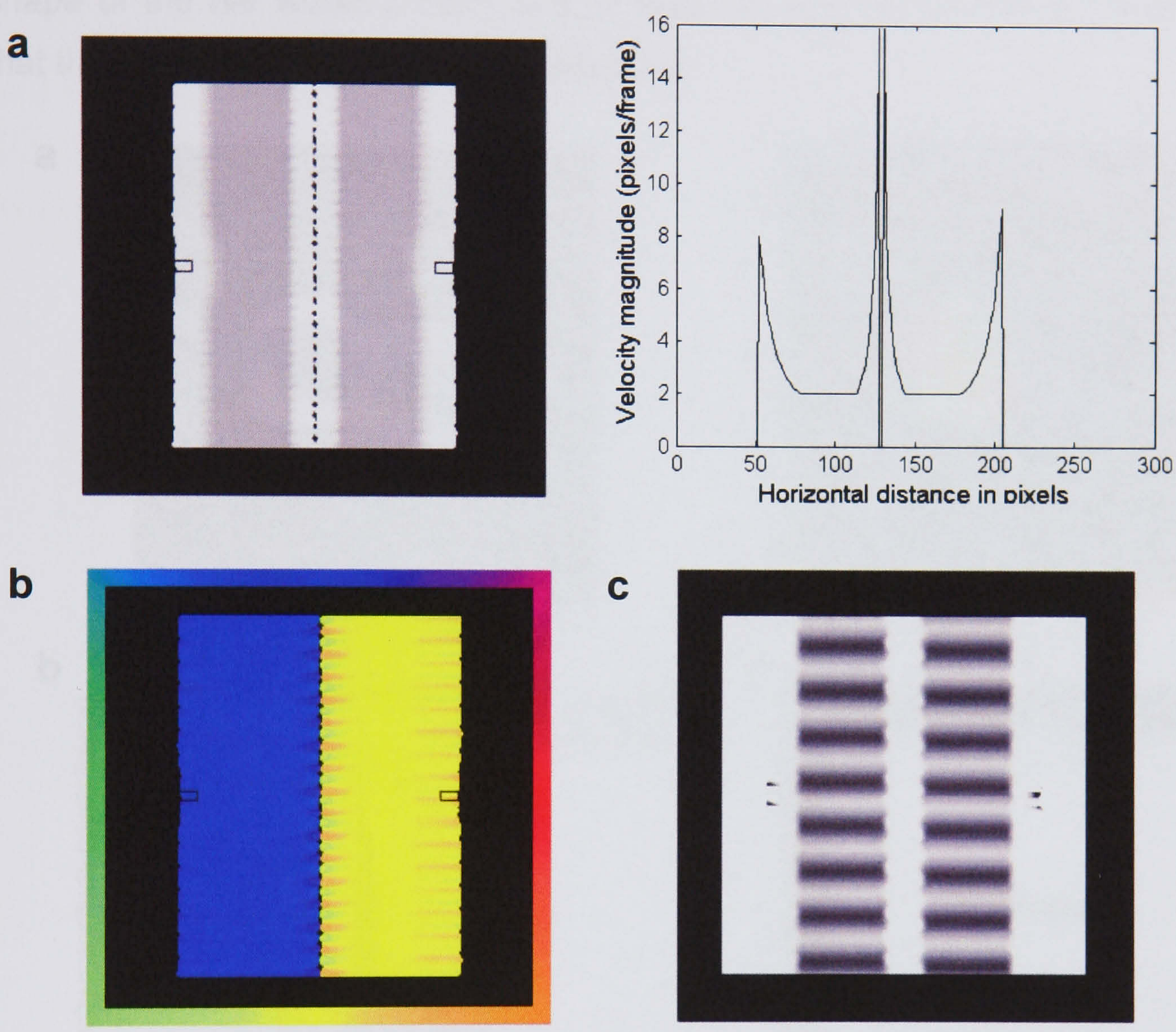


Fig. 5.12. Inputting two oppositely moving gratings (left upwards, right downwards), with flashes either side into the 2nd model with uniform averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 31×31 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude output for the output frame in which the value of the flash peaks. Thresholded for values over 3. (0-3: black – white). Plotted is the velocity magnitude along the horizontal line through the middle of the image. (b) Direction of motion for the output frame. (c) Reconstructed image corresponding to the velocity output.

It becomes apparent with the rotating bar stimulus that this version of the model does not produce suitable results for discrete motion. The averaging has some

non-intuitive effects on the velocity field of the bar, see Fig. 5.13. The velocity of the bar separates into two areas, separated by high velocity edges with zero velocity in-between them. This is no longer consistent with the velocity that the shape of the bar would predict and as such we can see in the reconstruction that the shape of the bar is greatly distorted.

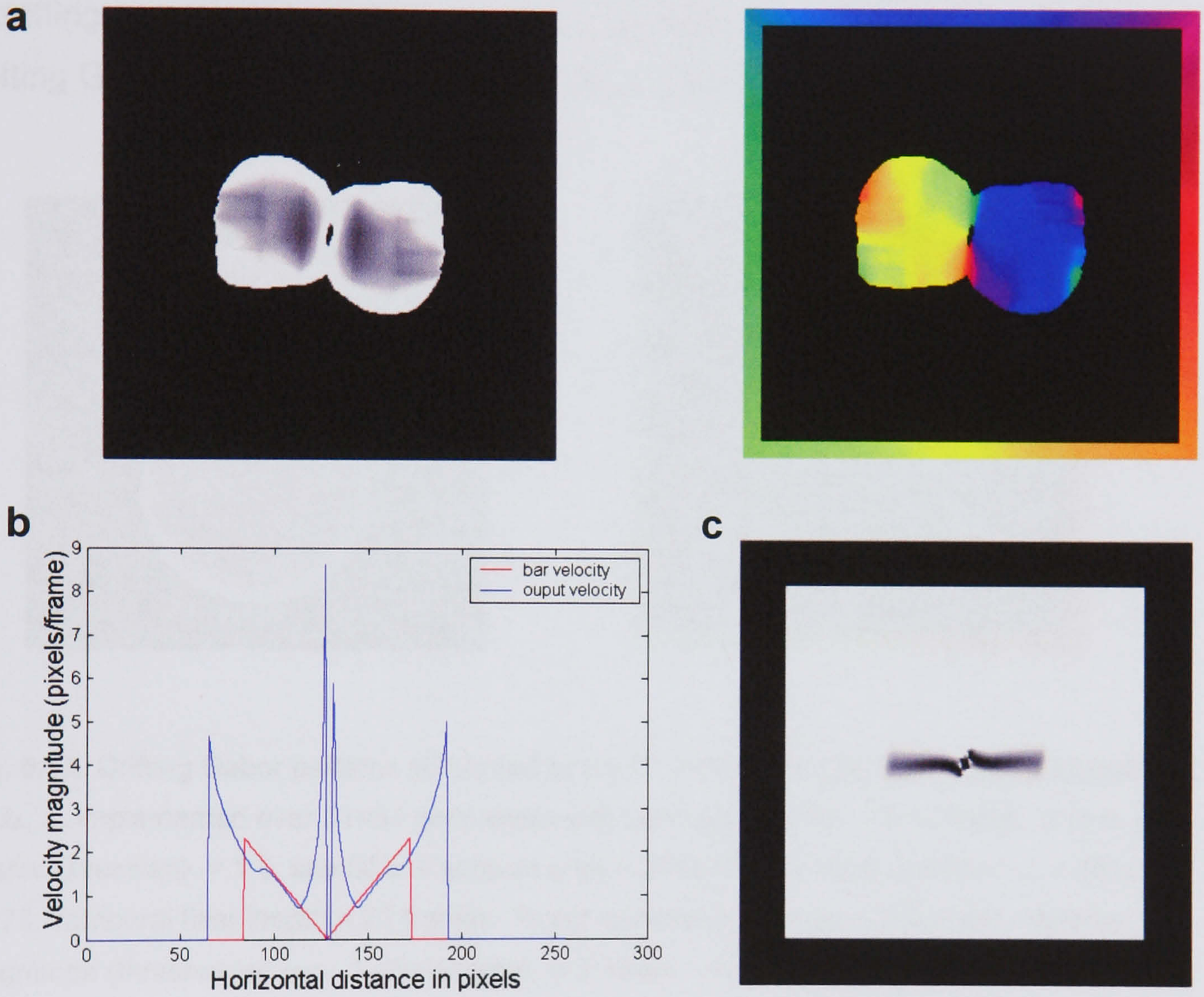


Fig. 5.13 Inputting a sequence of an anticlockwise rotating bar. Output from the 2nd model with uniform averaging over $\hat{s}_{\|}$, $\check{s}_{\|}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 31×31 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude, (thresholded at 3, (0-3: black – white)) and direction output. (b) Velocity magnitude plotted for the mid-horizontal line of the velocity output and the corresponding true velocity of a rotating bar at the horizontal position. (c) Reconstruction corresponding to the velocity frame.

It appears that averaging can reduce the motion magnitude measured around the flash and does extend the motion field beyond the boundaries of the moving

object. However, we need to find a method that does not introduce these high velocity artefacts.

To try to counteract these edge effects, we considered a different method of averaging. Instead of a uniform filter, a 2D Gaussian was used to smooth the four different speed measures. This strategy still gave us the correct speed for a drifting sine grating (correct to 4 s.f.), and the area of high velocity around a drifting Gabor patches is less, but similar problems remain (Fig. 5.14).

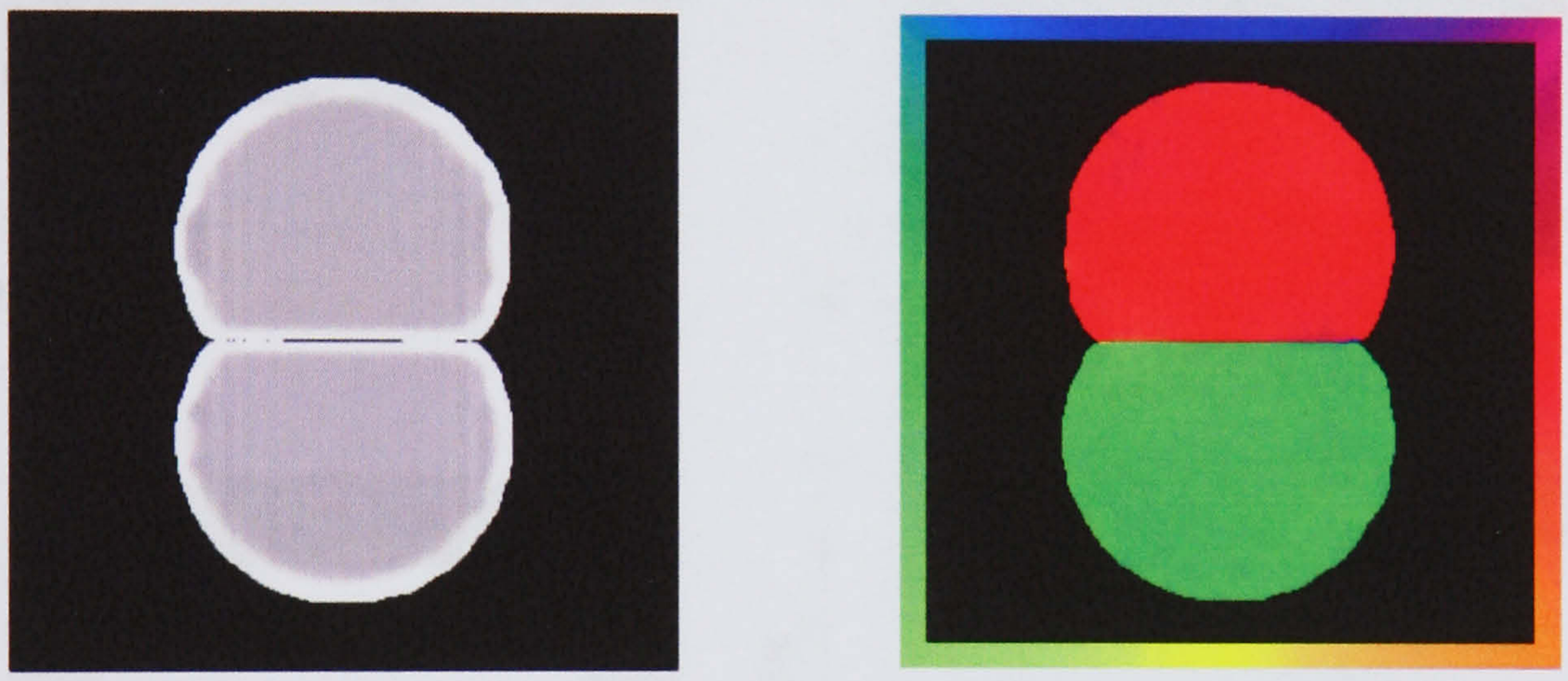
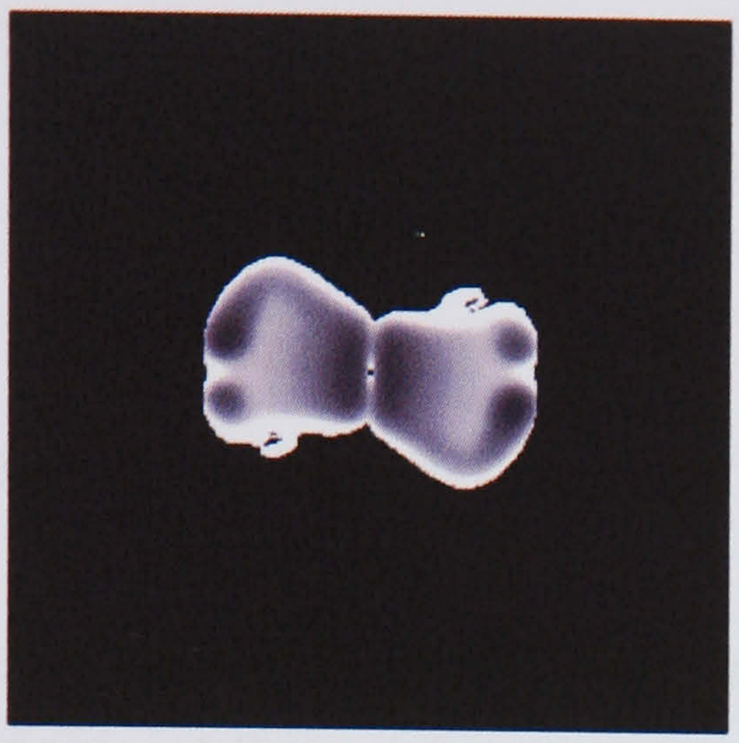
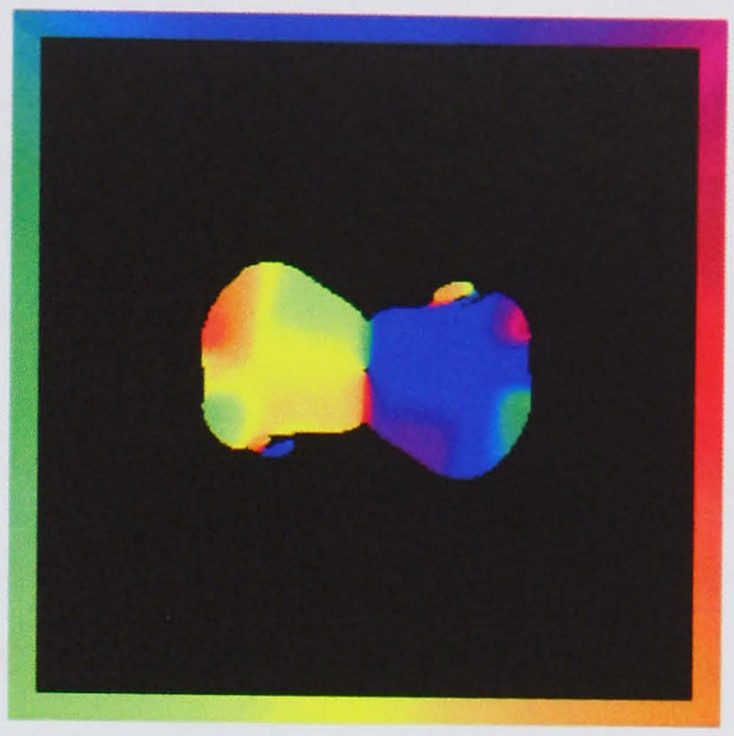
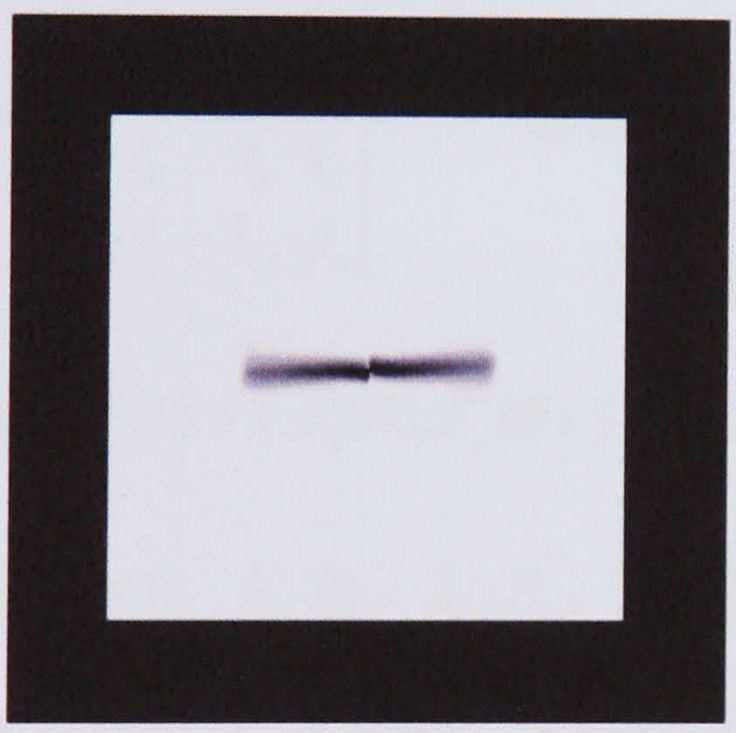


Fig. 5.14. Drifting Gabor patterns presented to the 3rd model with Gaussian averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 31×31 pixel areas with Gaussian s.d = 5. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. Velocity magnitude (thresholded over 3 pixels/frame, 0-3: black – white) and direction.

With a larger blur filter, as with the uniform averaging, the extent of the high velocity edges increases. So a rotating bar is still distorted. As we can see in Fig. 5.15, although the velocity profile changes more smoothly from end to end there is still a disruptive effect in the spatial representation at the middle of the bar. We now consider a different level at which to implement the velocity averaging.







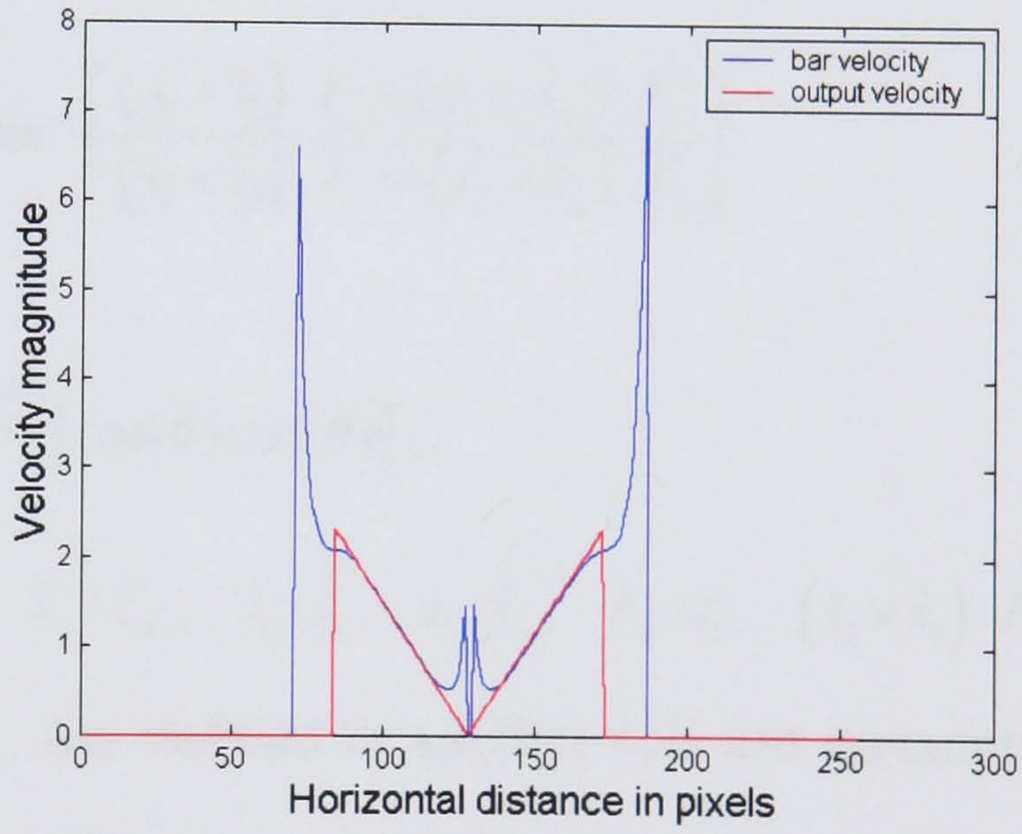


Fig. 5.15 Output frames for a sequence of an anticlockwise rotating bar. The 3rd model with Gaussian averaging over $\hat{s}_{\parallel}$, $\check{s}_{\parallel}$, $\hat{s}_{\perp}$, $\hat{s}_{\perp}$ implemented over 51×51 pixel areas with Gaussian s.d.=10. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) One frame of velocity output magnitude (thresholded above 3 pixels/frame, 0-3: black-white) and direction. (b) Corresponding reconstructed image. (c) A plot of the velocity output across the bar when horizontal shown with the true bar velocity. We see a similar velocity pattern as before.

5.3.2 Pooling over the components of the ratio operation (Version 4 & 5)

Project the components of $\hat{\boldsymbol{s}}, \check{\boldsymbol{s}}$ onto sine and cosine functions to extract the fundamental Fourier coefficients. Produce dot products of $\hat{\boldsymbol{s}}, \check{\boldsymbol{s}}$ with

$$\boldsymbol{F}(\theta) = (\ F_{\parallel}(\theta), F_{\perp}(\theta)\ ) = \sqrt{2/m}\left[cos(\theta), \sin(\theta)\right]$$

Instead of implementing averaging over the speed and inverse speed matrices, we now consider the effect of averaging each dot product element of the determinants in the quotient used for velocity calculation and each of the elements of the quotient used for the direction calculation. So each of the dot products involved in the calculations of speed and direction as were shown in Eqn. 4.11 and 4.12 is first averaged, before computing the final result.

$$S^2 = \frac{\begin{vmatrix} \hat{s}_{\parallel} \cdot F_{\parallel} & \hat{s}_{\parallel} \cdot F_{\perp} \\ \hat{s}_{\perp} \cdot F_{\parallel} & \hat{s}_{\perp} \cdot F_{\perp} \end{vmatrix}}{\begin{vmatrix} \hat{s}_{\parallel} \cdot \check{s}_{\parallel} & \hat{s}_{\parallel} \cdot \check{s}_{\perp} \\ \hat{s}_{\perp} \cdot \check{s}_{\parallel} & \hat{s}_{\perp} \cdot \check{s}_{\perp} \end{vmatrix}} \qquad \text{direction} = \tan^{-1}\left(\frac{\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\parallel} - \left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\perp}}{\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\perp} + \left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\parallel}}\right)$$

Where $\boldsymbol{F}(\theta) = (\ F_{\parallel}(\theta), F_{\perp}(\theta)\ ) = \sqrt{2/m}\left[cos(\theta), \sin(\theta)\right]$.

The terms $\hat{s}_{\parallel} \cdot F_{\parallel}$, $\hat{s}_{\parallel} \cdot F_{\perp}$, $\hat{s}_{\perp} \cdot F_{\parallel}$, $\hat{s}_{\perp} \cdot F_{\perp}$, $\hat{s}_{\parallel} \cdot \check{s}_{\parallel}$, $\hat{s}_{\parallel} \cdot \check{s}_{\perp}$, $\hat{s}_{\perp} \cdot \check{s}_{\parallel}$, $\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\perp}$, $\left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\parallel}$, $\left(\check{s}_{\parallel} \times \hat{s}_{\parallel}\right) \cdot F_{\parallel}$, $\left(\check{s}_{\perp} \times \hat{s}_{\perp}\right) \cdot F_{\perp}$ (as defined in section 4.2) are calculated for each pixel and so form image sized matrices that can be spatially averaged as before. This means averaging occurs after the projection onto sine and cosine basis functions.

The uniform filter is implemented first. Again, it was verified that this gives the correct results for a simple drifting sine grating (correct to 4 s.f). Looking at the velocity output for the drifting Gabor patches (Fig. 5.16), we can see that this method using the same size of averaging filters extends the effect of motion further out as well as smoothing it.

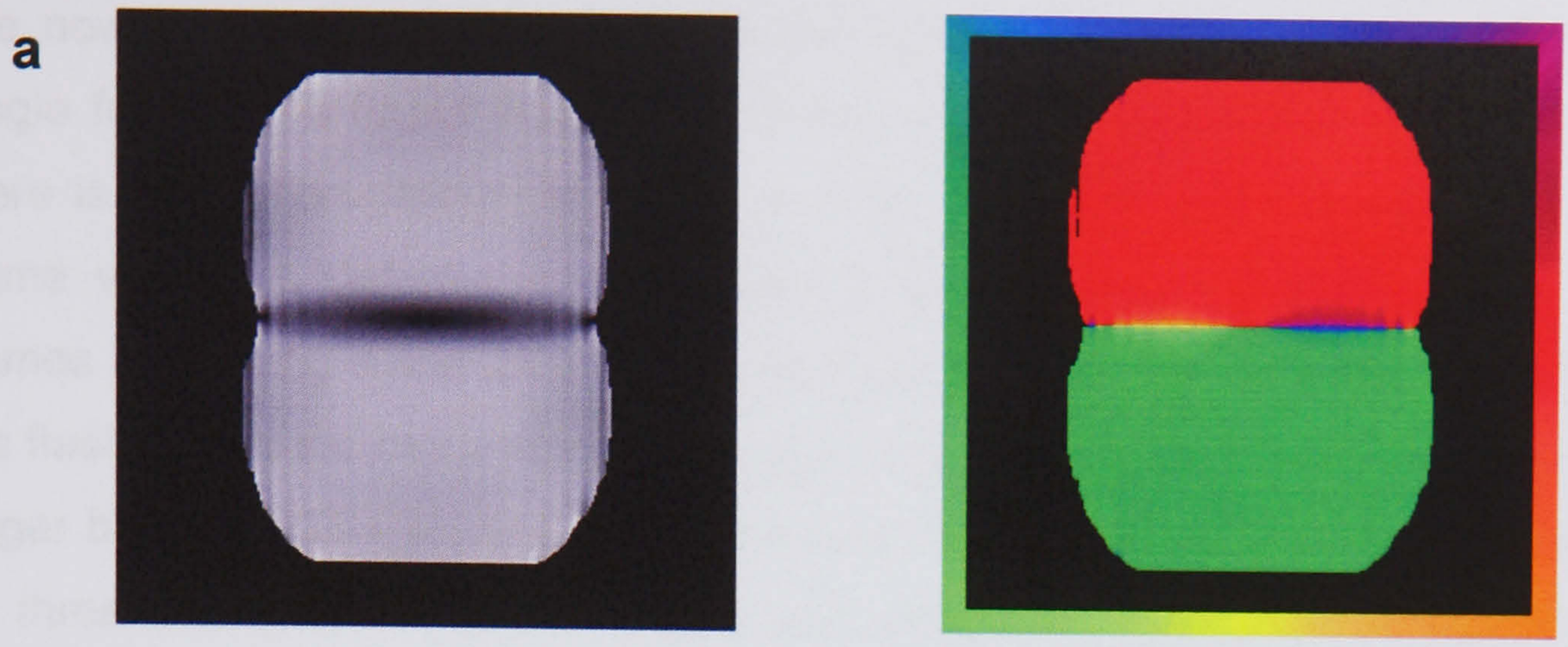
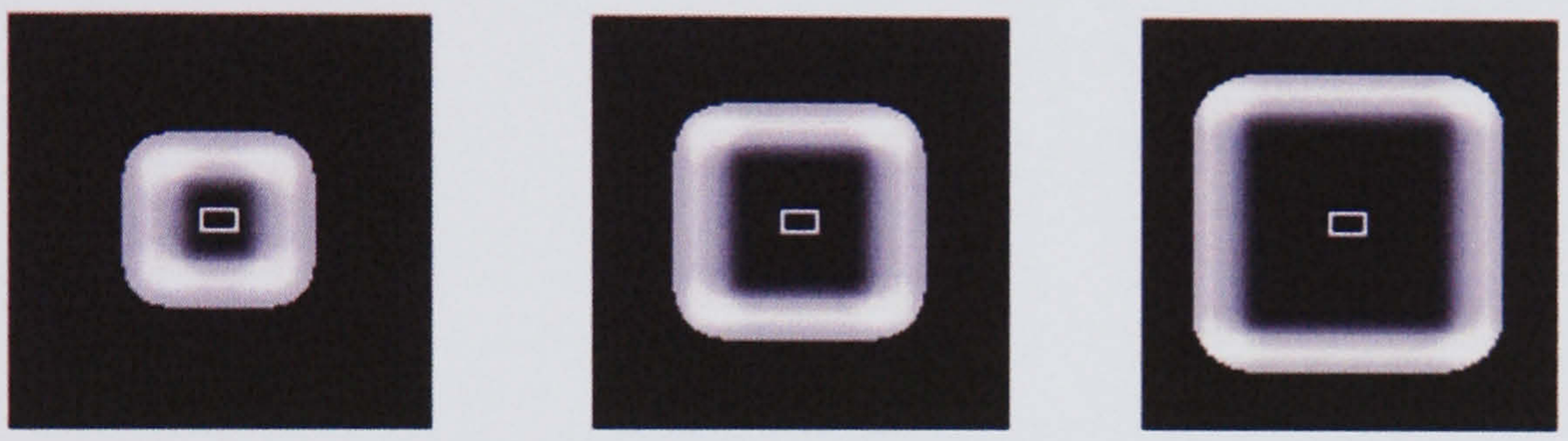


Fig. 5.16 A sequence of drifting Gabor patches as input to the 4th version of the model with uniform averaging over the elements of the motion quotients over 31×31 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23x23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Single frame from velocity output magnitude (0-3:black – white, max = 2.8 pixels/frame) and direction. (b) Plot of velocity magnitude along a horizontal line through the middle of the top Gabor patch from the output frame in (a). (c) Reconstructed version of the image, shown with plot of horizontal line through the middle of each of the Gabor patches.

We now take a look at this model's response to single flash presented in a single frame. It is found that for the three motion window sizes 31, 51 and 71 there is no motion present at the flash (Fig. 5.17). The velocity present in the frame varies as before during the spatial establishment of the flash. In the frames containing some motion the results are spread out in a ring away from the flash. The zone becomes increasingly narrow and further from the flash with larger blur size. The square, discrete look to this motion band is again caused by thresholding in the motion output and directional illustration as mentioned above. However, the presence of this band varies as before and in frame 10, in which the spatial response to the flash peaks, there is no motion present.



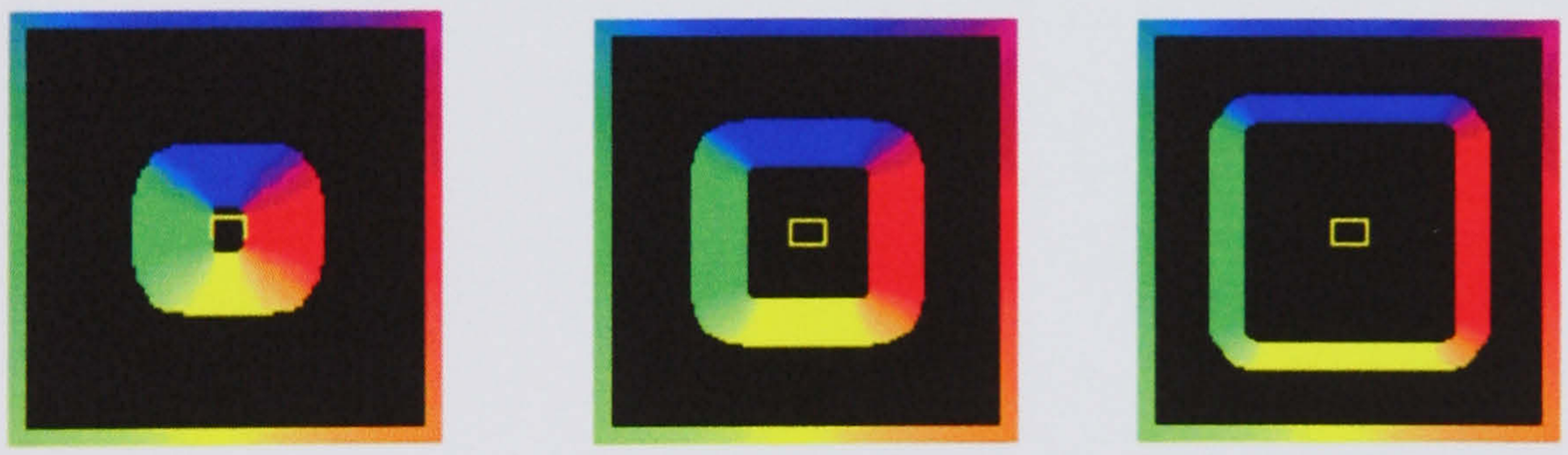


Fig. 5.17 Single frame of velocity output (magnitude and direction) for velocity blur size of 31×31, 51×51 and 71×71 pixels. Results for 4th version of model. Maximum velocity = 6.7 pixels/frame for each blur size. Note: velocity for frame 11 after flash presentation shown in which there is velocity present. Velocity magnitude is thresholded over 3 pixels/frame, 0-3:black-white. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels.

If we examine the output of the drifting gratings with flashes presented horizontally either side, we see the effect of the flashes on the motion field becomes less (Fig. 5.18). With the greater spreading out of the motion field, we

also see a smearing out of values, so that the velocity output becomes less towards the edge of the motion field. Because there are no high areas of velocity around the flashes in this version we see no pixels taking inconsistent luminance values around the reconstruction of the flashes.

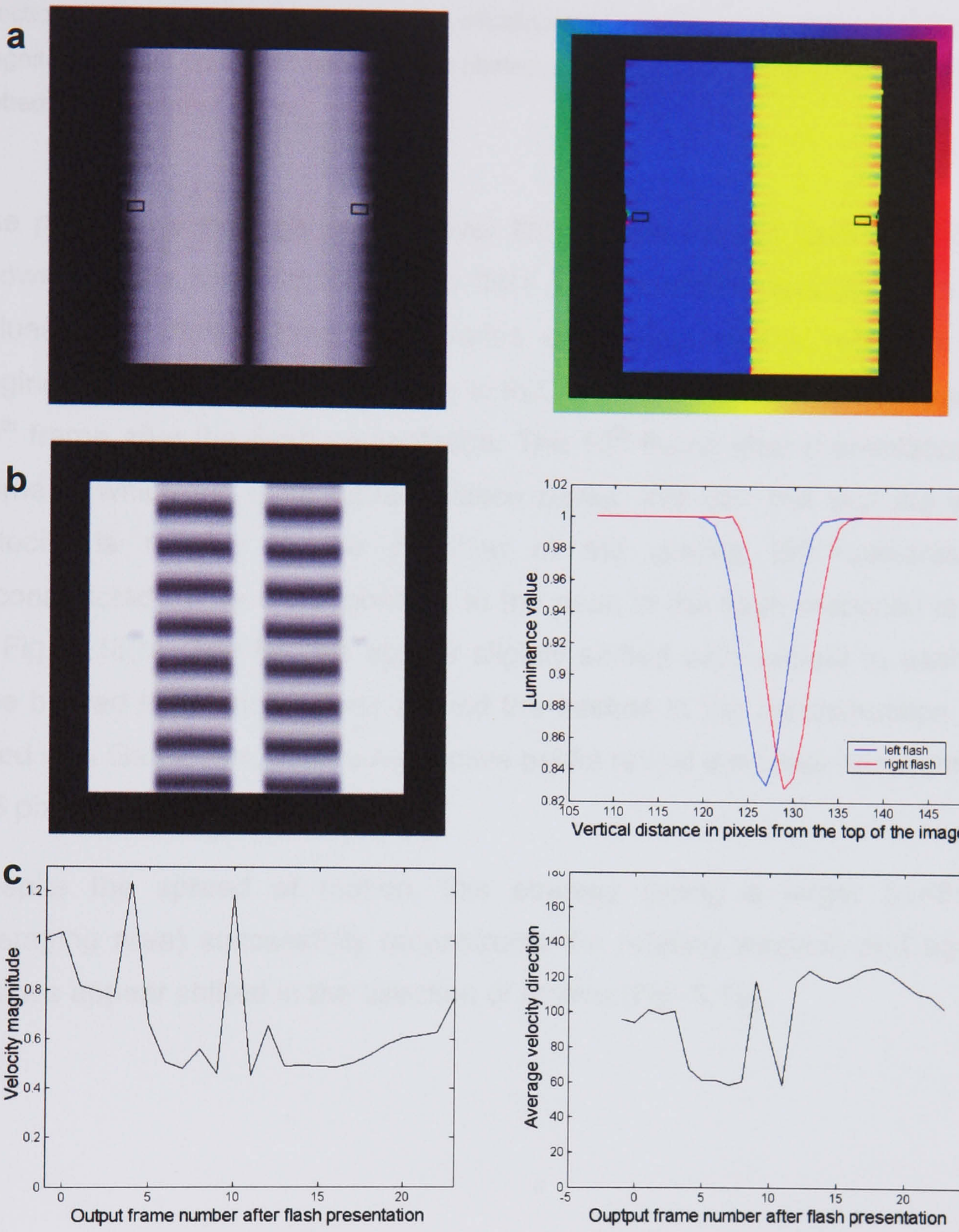


Fig. 5.18 Presenting two oppositely moving gratings, with flashes either side to the 4th model with uniform averaging over the elements of the motion quotients implemented over 31×31 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max = 2 pixels/frame) and direction output for the output frame in which the value of the flash peaks in the temporal blur. Values scaled 0-black, 3-white. (b) Reconstructed image corresponding to the two velocity outputs. Shown with plots of the luminance values along the vertical lines through the middle of each of the two flashes in the reconstruction. The left flash is higher in the image than the right flash, consistent with the

direction of the grating's motion. (2.6 pixel misalignment found). (c) The average velocity magnitude plotted on the left and direction plotted on the right over the area of the left flash plotted at each output frame.

The plot of the average velocity over the area of the left flash (Fig. 5.18(c)), shows that the presentation of the flash initially slightly reduces local motion values. The motion magnitude varies over time, with a reduction at the beginning of the temporal response to the flash and a peak in the 5th frame and 10th frame after the flash presentation. The 10th frame after presentation is the frame in which the flash representation peaks. We can see that the average velocity is roughly in the direction of the grating (90°=upwards). The reconstructed frame corresponding to the peak in the flash response is shown in Fig. 5.18(b). The flashes appear slightly shifted with respect to each other. The blurred luminance values around the flashes in the reconstruction can be fitted with Gaussians, whose respective peaks reveal a relative misalignment of 2.6 pixels.

Despite the spread of motion, this strategy (using a larger 51×51 pixel averaging area) successfully reconstructs the rotating windmill and again the flashes appear shifted in the direction of motion (Fig. 5.19).

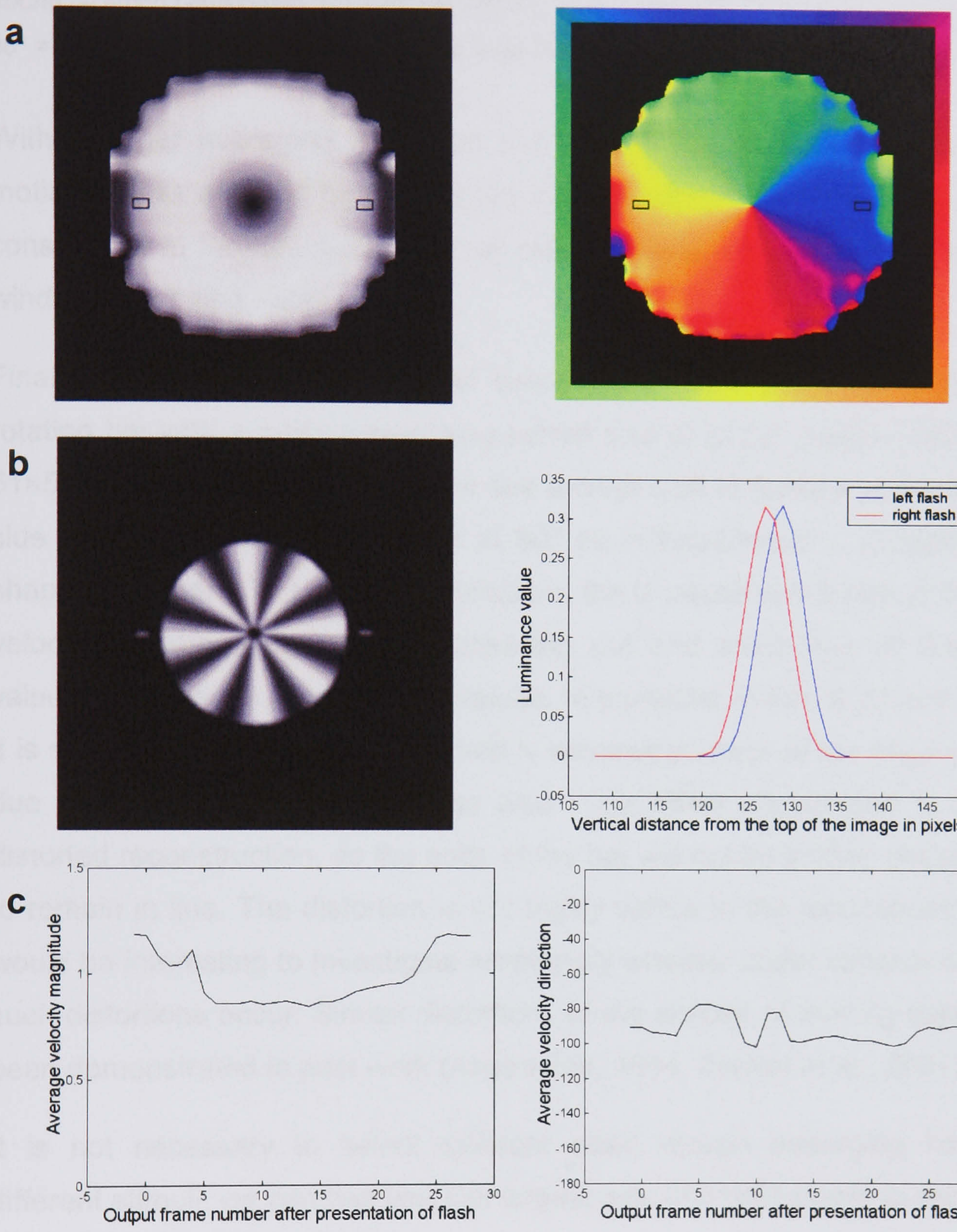


Fig. 5.19 An anticlockwise rotating windmill with horizontal flashes either side processed by the 4th model with uniform averaging over the elements of the motion quotients implemented over 51×51 pixel areas. Parameters: σ (s.d. of spatial Gaussian) = 1.5, spatial blur support area = 23×23 pixels, temporal filter parameters: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Single frame from velocity output, magnitude (max 1.3 pixels/frame -white, 0-black) and direction. (b) Reconstructed version of image, shown with the plot of the vertical line of luminance values through the middle of each of the flashes. They are shifted relative to each other in the direction of motion by 1.7 pixels

(found by fitting Gaussians). (c) Average velocity magnitude (on left) and direction (on right) (-90° = downwards) at each frame over the area of the left flash.

With a larger averaging kernel on the velocity values, we can see that the motion is less affected by the flashes and the direction at the flashes is more consistently in the direction of the windmill (downwards for the left flash as the windmill is rotating anticlockwise).

Finally, it is shown that this model successfully reconstructs the shape of the rotating bar with a motion averaging kernel size of 31×31 pixels. With a larger 51×51 pixel averaging area we can see a clear shift in flashes presented either side of a bar moving at 3°/frame at 60° as in Experiment 1, Chapter 2. The shape of the bar is not distorted. However, the increase in the size of the spatial velocity averaging causes the spreading out and smoothing of the velocity values and delivers lower motion values. In particular in Fig. 5.20 and Fig. 5.21 it is shown that the motion calculated is somewhat lower at the edge of the bar due to this averaging over a large area. This effect should lead to a slightly distorted reconstruction, as the ends of the bar will not be shifted ahead enough to remain in line. The distortion is not highly visible in the reconstruction, but it would be interesting to investigate empirically whether under suitable conditions such distortions occur. Similar distortions of the shapes of moving objects have been demonstrated in past work (Ansbacher, 1944; Zanker et al., 2001).

It is not necessary to select different sized motion averaging kernels for different stimuli, we can just use this largest size of 51×51 pixels to recreate the effects in all the stimuli presented so far.

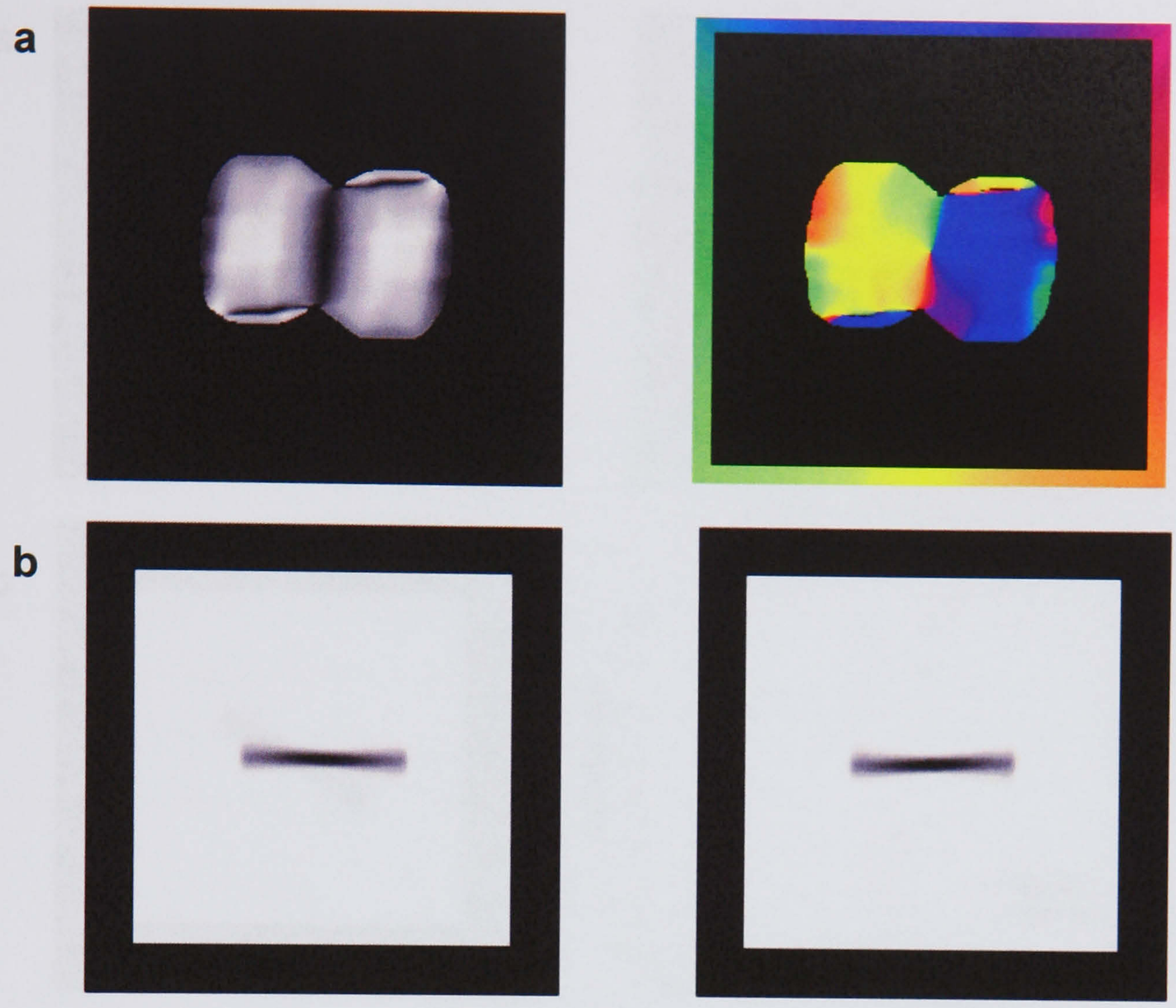


Fig. 5.20 Results for a sequence showing an anticlockwise rotating bar (3°/frame) for the 4th version of the model with uniform averaging over the elements of the motion quotients implemented over a 31×31 pixel window. Parameters: σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max: 1.6 pixels/frame - white, 0 - black) and direction for the output frame corresponding to the peak of the temporal response to the flash. (b) Corresponding blurred image and reconstruction. (c) Plot of the velocity output versus input for the rotating bar when horizontal, along a line of values through the horizontal.

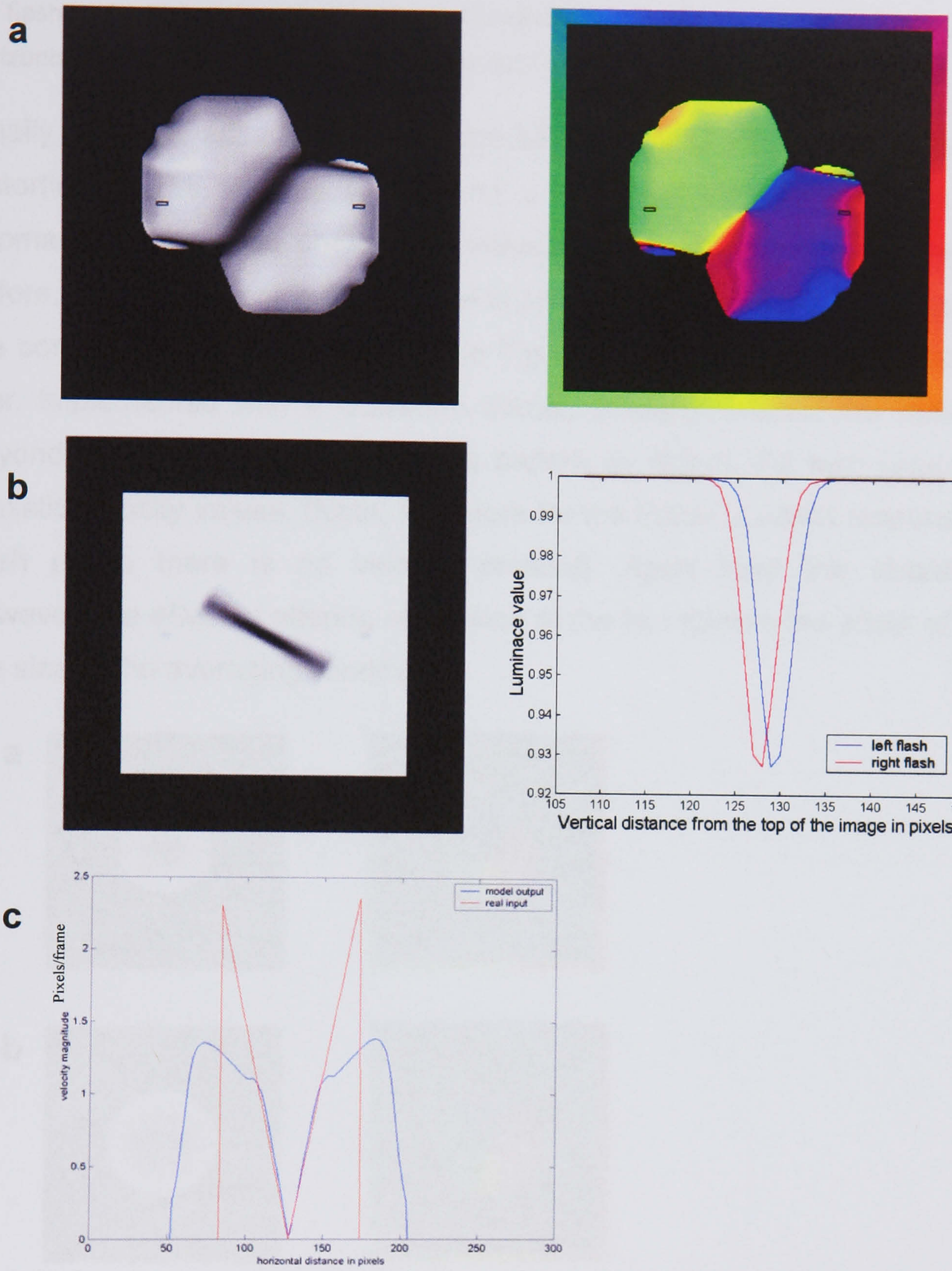


Fig. 5.21 Results for a sequence showing an anticlockwise rotating bar (3°/frame) with horizontal flashes occurring at the bar position 60° past the vertical. 4th model with uniform averaging over the elements of the motion quotients implemented over 51×51. Parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23x23 pixels, temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max=1.6 pixels/frame – white, 0 – black) and direction for the output frame corresponding to the peak of the temporal response to the flash. (b) Corresponding reconstructed image with a plot of the values along vertical lines through the middle of each of

the flashes. (c) Plot of the velocity values of a rotating bar along the line through at the mid-horizontal, input motion values versus model output.

Finally, one last adjustment to the model is made to ensure that there is less distortion of the velocity output and a smoother velocity profile with less anomalies at the edge of motion. Instead of the uniform smoothing kernel, as before, a Gaussian smoothing kernel is applied to the velocity elements used in the computation. As demonstrated in Fig. 5.22 and Fig. 5.23, a similarly large blur, implemented with a Gaussian kernel, similarly extends the motion field beyond the boundary of the moving pattern or object, but also returns more realistic velocity values. (Note, as before for the frame in which response to the flash peaks there is no velocity present). Apart from this improvement, however, the effect of altering of the size of the blur mirrors the effect of altering the size of the averaging window.

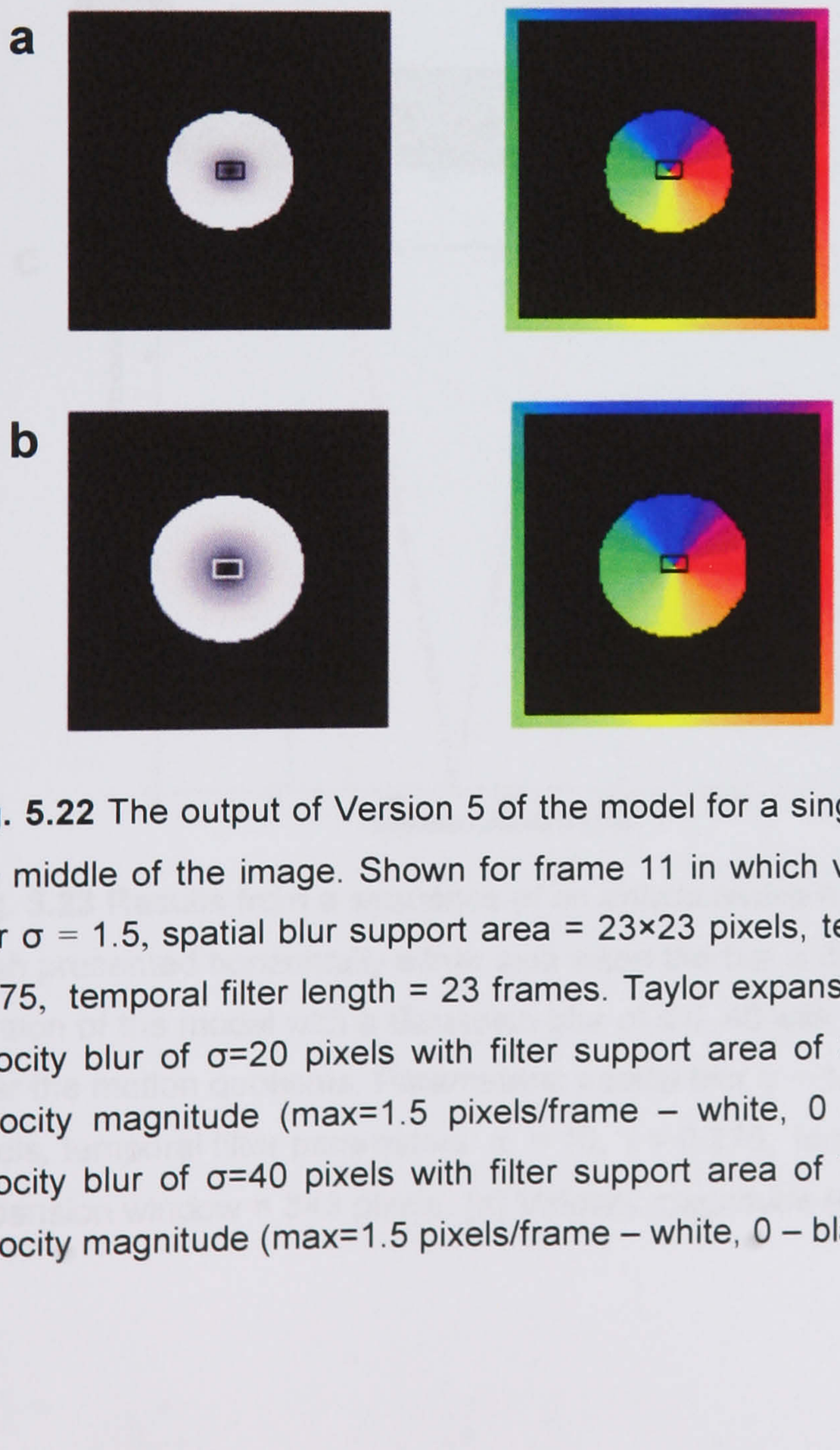


Fig. 5.22 The output of Version 5 of the model for a single flash presented in a single frame in the middle of the image. Shown for frame 11 in which velocity is present. Parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels, temporal filter parameters: $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Gaussian velocity blur of σ=20 pixels with filter support area of 35 pixels applied to motion quotients. Velocity magnitude (max=1.5 pixels/frame – white, 0 – black) and direction. (b) Gaussian velocity blur of σ=40 pixels with filter support area of 71 pixels applied to motion quotients. Velocity magnitude (max=1.5 pixels/frame – white, 0 – black) and direction.

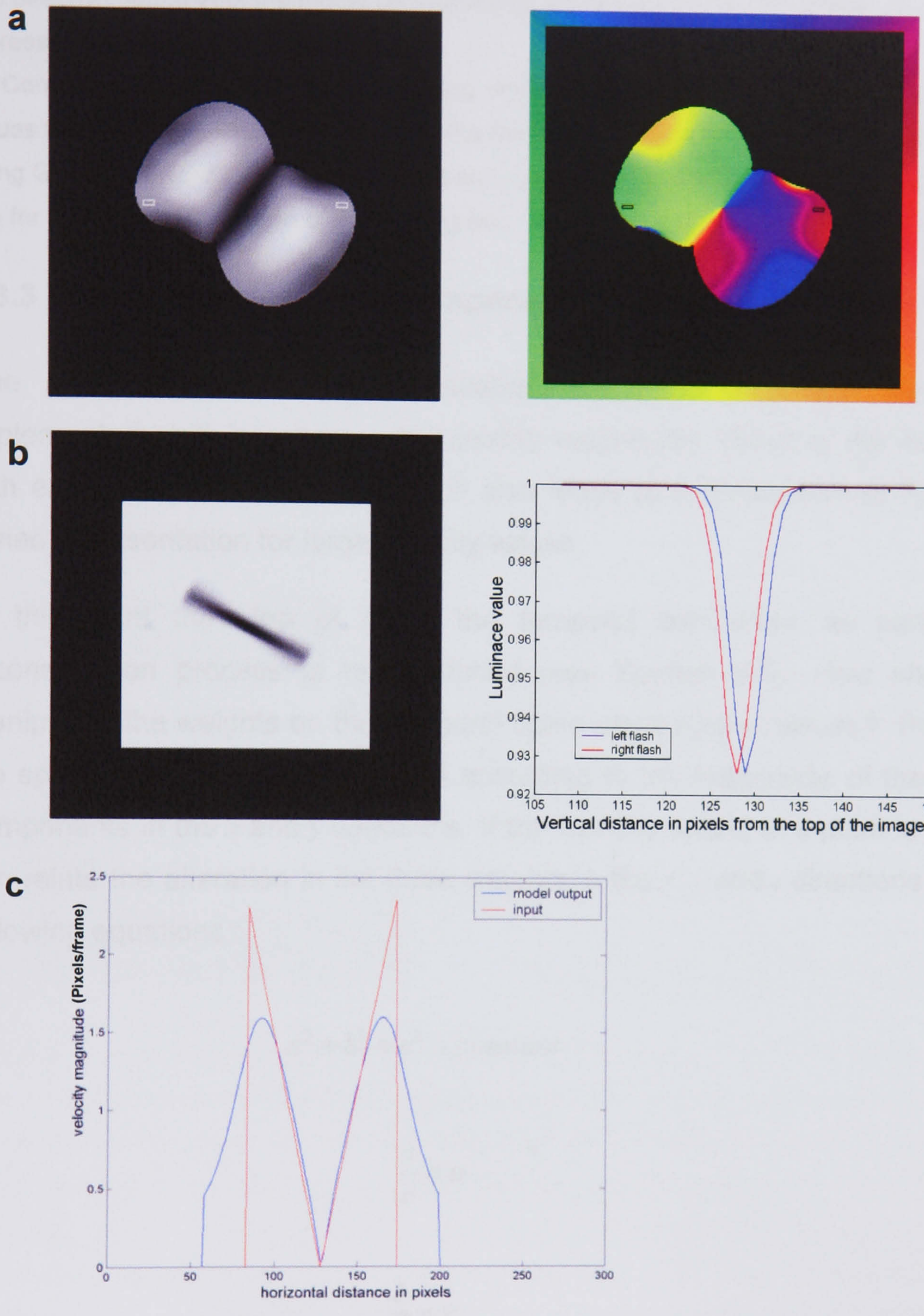


Fig. 5.23 Results from a sequence of an anticlockwise rotating bar (3°/frame) as before with a flash presented horizontally either side when the bar is 60° past the vertical. Using the 5th version of the model with a Gaussian blur of s.d. 40 with a 71×71 pixel support area applied over the motion quotients. Parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels, temporal filter parameters: $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max 1.6 pixels/frame - white, 0 -black)

and direction output from the model corresponding to the frame in which the flash representation peaks.
(b) Corresponding reconstructed image using motion feedback, shown with a plot of luminance values for the vertical line through each of the flashes. Misalignment of 1.1 pixels found by fitting Gaussians. (c) Plot of velocity output and real velocity magnitude along the mid-horizontal line for the horizontal position of the rotating bar.

5.3.3 Using a Taylor series in space and time

One problem with this simple spatial reconstruction is that clearly the implemented shift increases with velocity magnitude, which is not consistent with experimental results and which also leads to a breakdown in the Taylor series representation for large velocity values.

At this point the idea of using the temporal derivatives as part of the reconstruction process is re-introduced (see Section 4.5). How should we manipulate the weights on the temporal filters using motion values? Previously the spatial filters were manipulated according to the magnitude of the velocity components in the x and y directions. If the velocity output at a point is (u,v), we can relate the alteration in the three weights in the x, y and t directions with the following equations :

$$a^2 + b^2 + c^2 = constant$$

$$\frac{a}{c} = u \tag{5.1}$$

$$\frac{b}{c} = v$$

These weights can then be inserted in the 3D Taylor series representation

$$R_{i+p,j+q,k+r} = B_{i,j,k} + (p-a)\frac{\partial B_{i,j,k}}{\partial x} + (q-b)\frac{\partial B_{i,j,k}}{\partial y} + (r-c)\frac{\partial B_{i,j,k}}{\partial t} + \ldots \tag{5.2}$$

$R_{i,j,k}$ = pixel i,j in the kth frame of the rebuilt image R, , i in the x direction, j in the y direction

B = blurred image

These equations set an upper limit on the spatial shift, with the shift increasing in the temporal domain at higher speeds. The constant that sets the upper limit can be chosen arbitrarily and may be matched by data from experiments but, for now, the value of 1 will be chosen to see how this modification affects the pattern of results. In Fig. 5.24 we show the results from a sequence of a rotating bar with flashes presented either side, for the 5th version of the model, with this temporal reconstruction built in.

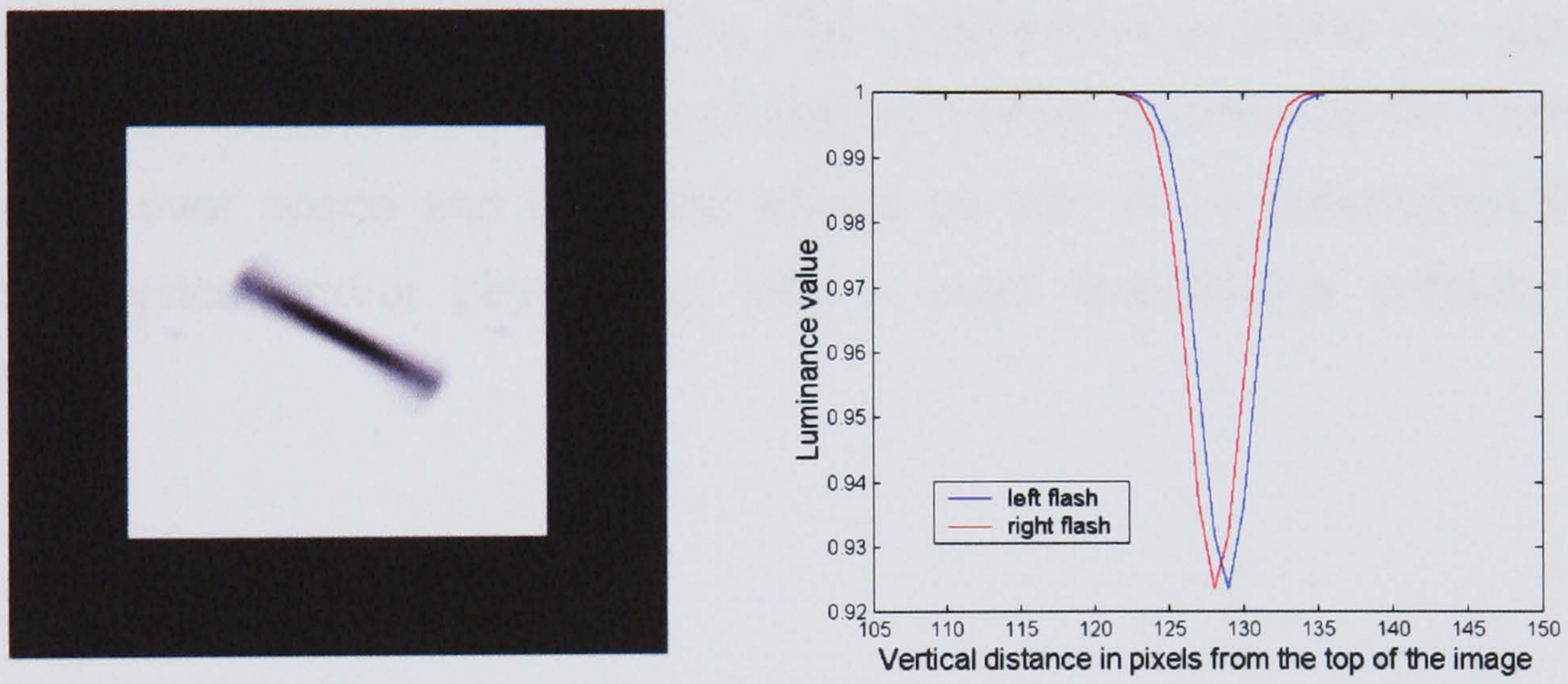


Fig. 5.24 Results from a sequence showing an anticlockwise rotating bar (3°/frame) as before with a flash presented horizontally on either side when the bar is 60° past the vertical. Using the 5th version of the model with temporal reconstruction. Every second frame is reconstructed using derivatives from the previous frame and the weights are given by Eqns. 5.3. With a Gaussian blur of s.d. 40 with a 71×71 pixel support area applied to the motion quotients. Parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. Misalignment 0.8 pixels, found by fitting Gaussians to the reconstructed values around the blurred flashes.

Version 5 of this model was successful in qualitatively reproducing the experimental results shown. We are going to use this model to analyse the

parameters, limits and uses of such a feedback mechanism. We will also consider the possibility of using reconstruction across time.

The parameters so far have been mostly chosen by the fact that they make the model work, i.e. they produce successful reconstructions of the sequences and reproduce the empirical results. As the model is qualitative at this stage, rather than quantitative, it is difficult to motivate these parameters based on physiological or psychophysical data. It seems fair to suggest that the spatial pooling area for the motion calculation should be larger than the spatial extent of the blurred derivative kernels, due to the larger size of MT receptive fields (Van Essen et al., 1981). At the same time this spatial extent needs to be limited by the size of the visual scene, which is linked with the extent over the visual scene of MT+ receptive fields. The constant that provides the upper limit in the temporal reconstruction is a similar parameter limiting how far motion can influence over space and time and should be able to be determined through psychophysics and/or physiology, with a more quantitative version of the model.

Chapter 6- Testing the model

The effects of varying the parameters of the McGM model for calculating motion have already been investigated in previous work (Dale, 2003; Johnston et al., 1999; Johnston et al., 1992). In this chapter I am going to examine the effects of the parameters introduced in the current version. Again, all the stimuli presented to the model consist of sequences of 256×256 pixel images.

The parameters of the spatial and temporal filters will be set at values determined in previous work (Johnston & Clifford, 1995; Johnston et al., 1999) (for a detailed description of the McGM see Chapter 4). This leaves us with three possible parameters to manipulate. These are: (1) the size of the window over which the approximations from the Taylor series are calculated (this has to be changed in line with the spatial blur applied by the model as explained in the previous chapter); and from the motion calculation we have (2) the size of the velocity blur (which we have already partly considered); and (3) the feedback parameter ξ. We will need to consider the robustness of these parameters and the range over which realistic results are produced. This range can help us make predictions about the biological mechanisms involved and predictions for further visual experiments. The final aim is to see what this model can tell us about the experiments described in the empirical section, how it behaves under the different paradigms and the implications for possible biological explanations.

6.1 - Investigating the model parameters

6.1.1 Velocity pooling window

First we consider the effect of changing the spatial extent of the velocity aggregation. We have a lower and an upper limit for this parameter, as the motion pooling window must be larger than that of the rebuilding window to represent the larger spatial extent of the motion sensitive cells, and smaller than the size of the input image. Some examples have already been shown in Chapter 5 of results with different sized motion averaging windows. It was found that for a drifting sine grating the accuracy of the velocity estimation was not altered by motion averaging, whatever the size of the window applied.

There are some technical limitations on the size of the motion aggregation window. Performing the convolution that pools over the motion calculation necessarily also shears half the filter support area size off the edges of the image. So the larger the filter applied, the less of the image is left as an output. Therefore, the motion aggregation area cannot be too large relative to the image. At its lower limit, when the standard deviation of the motion pooling window is very small, we end up with the first model, with no averaging.

For this parameter we need a value that results in the motion influence extending a suitable distance beyond the stimulus (capable of inducing a shift in a nearby object), without the motion aggregation window becoming too large. A value of 40 pixels will be used for the standard deviation of the Gaussian used to average over the components of the motion calculation.

6.1.2 Feedback parameter ξ

The size of any induced shift is linearly proportional to the feedback parameter ξ, up to the point where the shift becomes too large to implement the Taylor series reconstruction. A value of ξ=1 shifts the image ahead one frame in time (for a constant velocity input), and a value of ξ=k shifts the object ahead k

frames to the point where k becomes too large. However, multiplying the velocity input by a value greater than 1 highlights the mismatch between the smoothed motion and the real motion of the bar, causing the reconstructed image of the bar to become warped. The effect of using values of ξ=2 and ξ=4 in the reconstruction of a rotating bar is demonstrated in Fig.6.1.

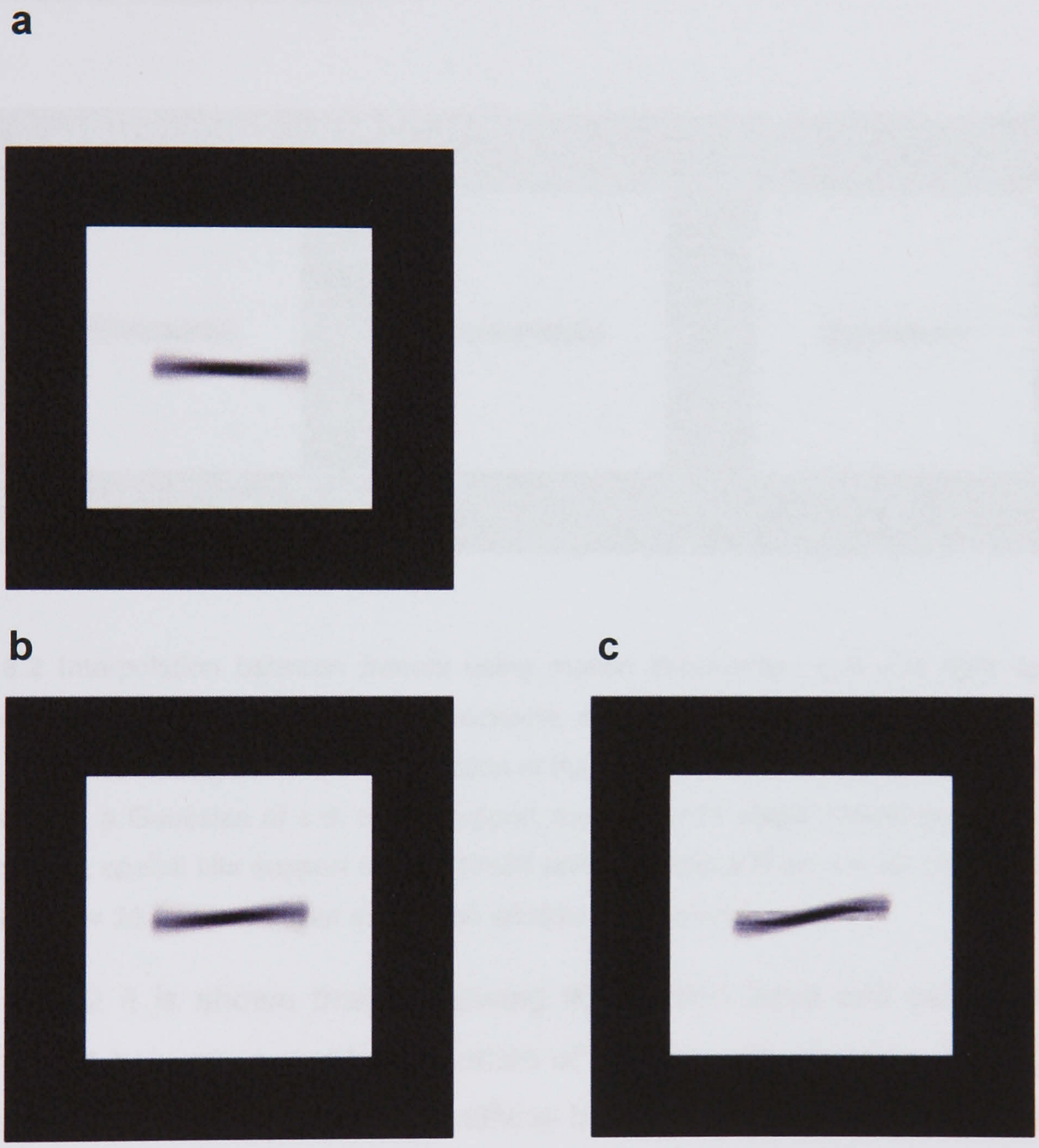


Fig. 6.1 Results from the sequence of an anticlockwise rotating bar produced by the 5th version of the model, ξ=2 and ξ=4, velocity is multiplied by 2 and 4 respectively before feeding into the representation. Motion quotient blurred with a Gaussian of σ = 40, support area = 71×71. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23x23 pixels; temporal filter: α = 10, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Single frame blurred in space-time (b) the corresponding reconstruction with ξ=2 (c) with ξ=4.

We could also use a value of $\xi < 1$, which would result in a reconstruction that lies between each frame of the output sequence produced by blurring the input sequence in space and time. Using $\xi < 1$ for reconstruction raises the possibility of using the motion field to produce smooth motion updates between samples. However, a reduced motion feedback parameter would also reduce the size of the effect of motion on position.

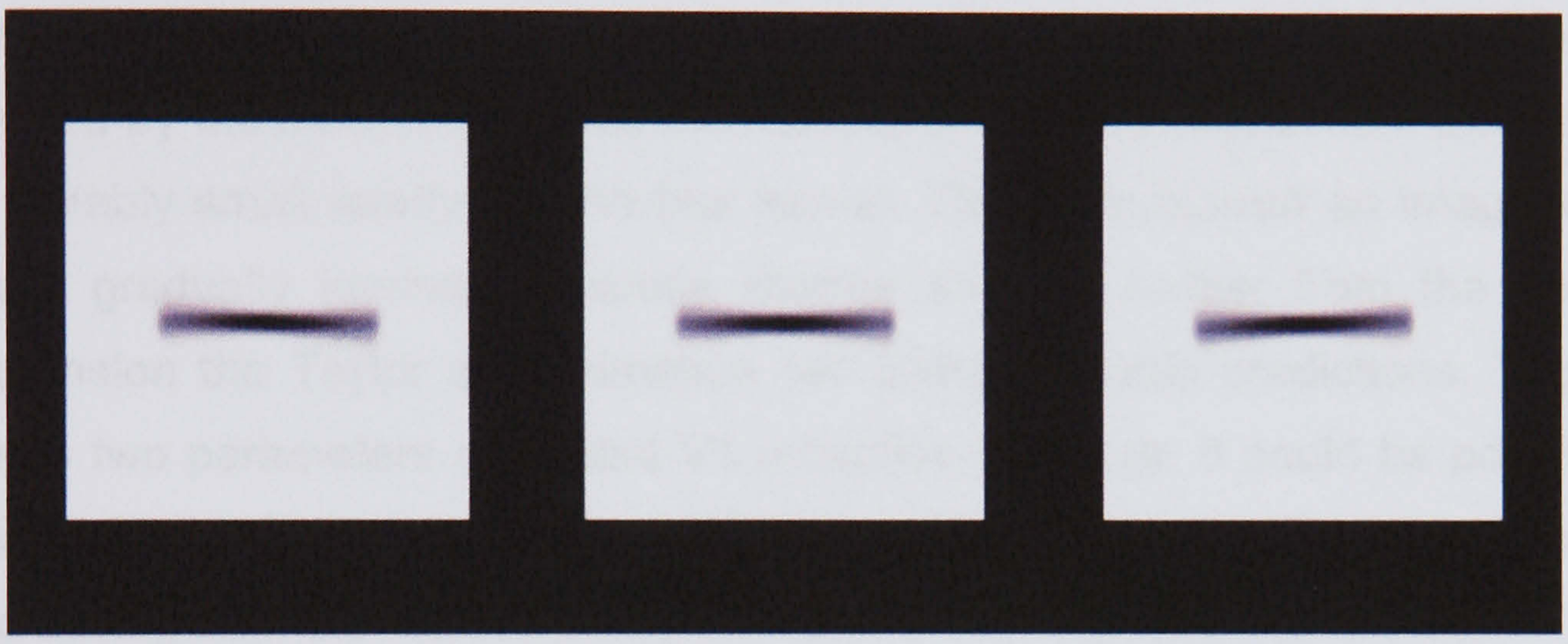


Fig. 6.2 Interpolation between frames using motion information. Left and right consecutive images from the sequence of an anticlockwise rotating bar are shown blurred in space and time. The middle image is the reconstruction of the image on the left with ξ=0.5. Motion quotient blurred with a Gaussian of s.d. of 40, support area = 71×71 pixels. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels.

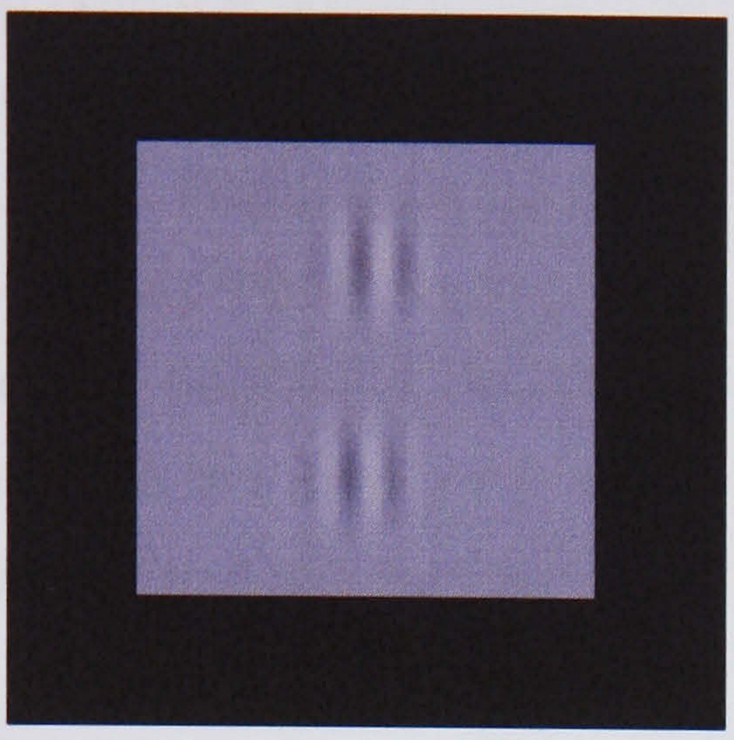
In Fig. 6.2 it is shown that by halving the motion input one can successfully interpolate between successive frames of the blurred sequence. The position of the bar in the reconstruction lies halfway between the two successive frames.

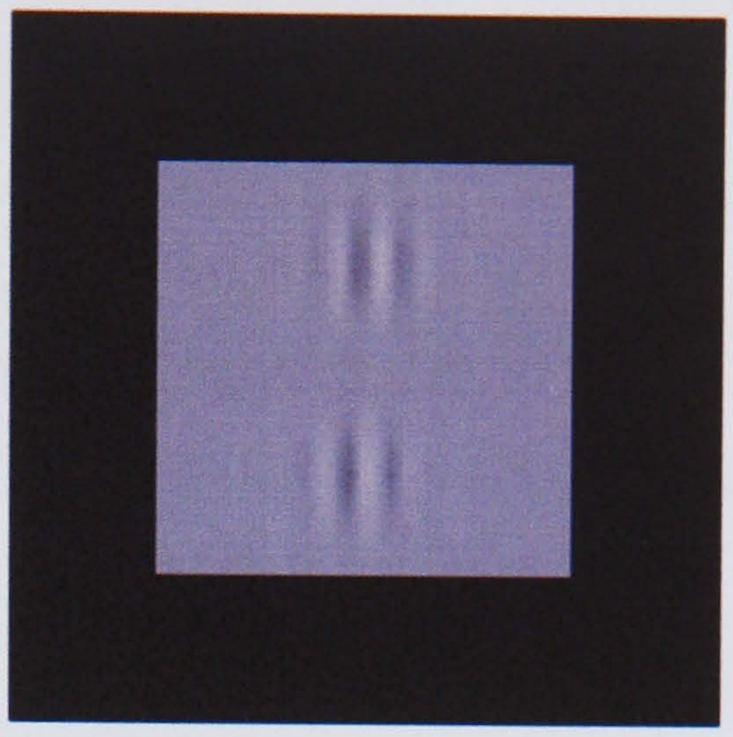
We can also consider the form of the velocity input function. According to empirical evidence (De Valois & De Valois, 1991; Whitney & Cavanagh, 2000) the size of the motion induced misalignment is not linearly proportional to the velocity magnitude. So far we have only considered a linear input, where ξ is a constant multiplier, rather than some other function of the velocity. In this situation the size of the shift will simply keep increasing with motion. However,

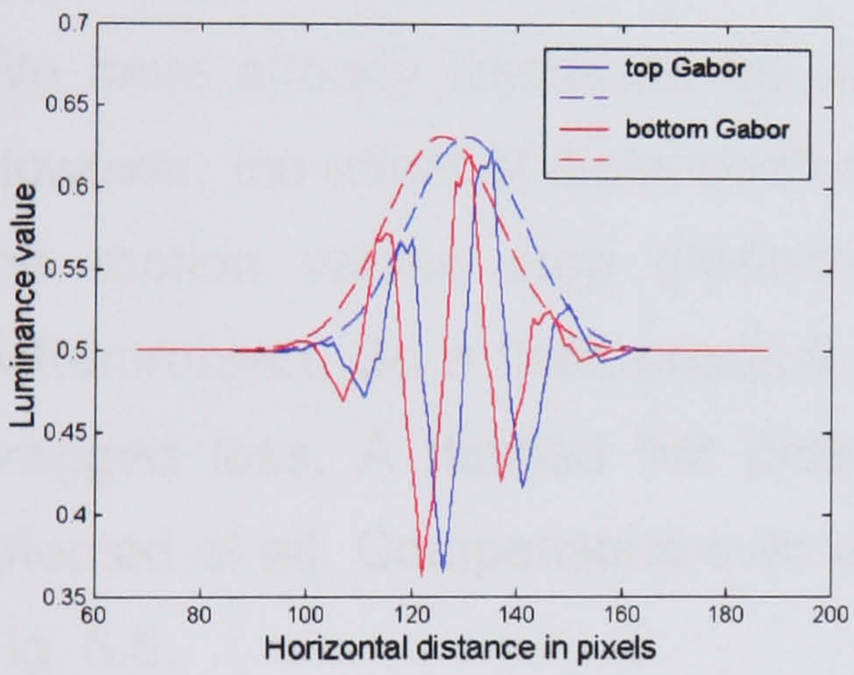
as mentioned above, a large shift cannot be implemented using the Taylor reconstruction algorithm. Temporal reconstruction has the advantage that the shift parameters are no longer linearly related to speed (see Eqns. 5.3).

6.1.3 Reconstruction expansion area

The third parameter introduced is the sampling rate and area for the Taylor series expansion. This is considered to be related to the size of the V1 receptive fields. This parameter has to be proportional to the original blur applied by the motion model as the reconstruction only succeeds if this window is suitably small relative to the blur kernel. The more blurred an image is, the more gradually luminance values change and the further from the point of expansion the Taylor approximation can deliver reliable predictions. Together these two parameters represent V1 receptive field size. It could be possible to approximate the effects of changes in receptive field size with eccentricity by increasing both the window size parameter and the spatial blur kernel.







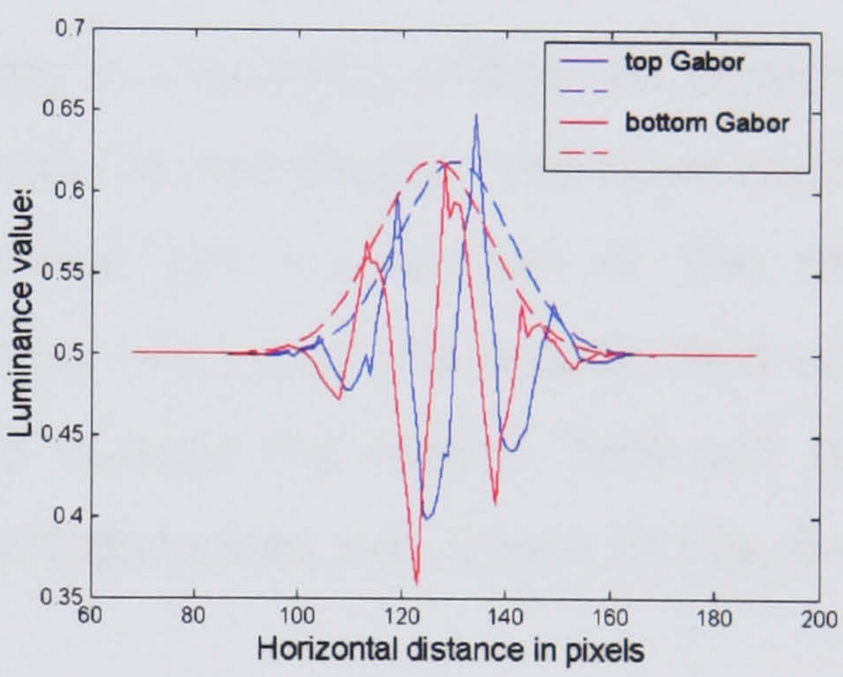


Fig. 6.3 Drifting Gabor (top - rightwards, bottom - leftwards) patches input into version 5 of the model. Velocity quotient blur parameters: σ= 40 pixels, support area= 71×71 pixels. (a) A reconstructed image from the sequence. Model parameters, spatial blur : σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter: α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (b) Same image reconstructed using model parameters: spatial blur σ = 2.5, spatial blur support area = 37×37 pixels; temporal filter : α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 5×5 pixels. (c) Plot of the luminance along horizontal lines through the middles of each of the Gabor patches in (a). Misalignment of 4.5 pixels found by fitting Gabor functions. (d) Plot of the luminance along horizontal lines through the middles of each of the Gabor patches in (b). Misalignment of 4.2 pixels found by fitting Gabor functions.

In Fig. 6.3 we compare results for the sequence of Gabor patches drifting in opposite directions. One of the frames from the sequence is reconstructed using both the original spatial blur of σ=1.5 and a reconstruction window of 3×3 pixels and a spatial blur of σ=2.5 with a reconstruction window of 5×5 pixels (Fig. 6.3). There is no great difference found between relative misalignments. This implies that changing receptive field size alone is not enough to model the effects of peripheral viewing on motion induced displacements. Other factors may come into play such as different retinal sampling size, different temporal resolution and different perceived velocities.

6.2 - Varying stimulus parameters

6.2.1 Distance from motion

We have already discussed the difficulty in modelling effects of eccentricity. However, the effect of distance from motion is reproduced. We have seen how the motion values drop gradually outside the boundaries of the moving pattern/object. So a flash presented further out along this motion field edge is dragged less. A flashed bar presented outside the motion field will not be affected at all. Comparisons over different distances are shown in Fig. 6.4 and Fig. 6.5.

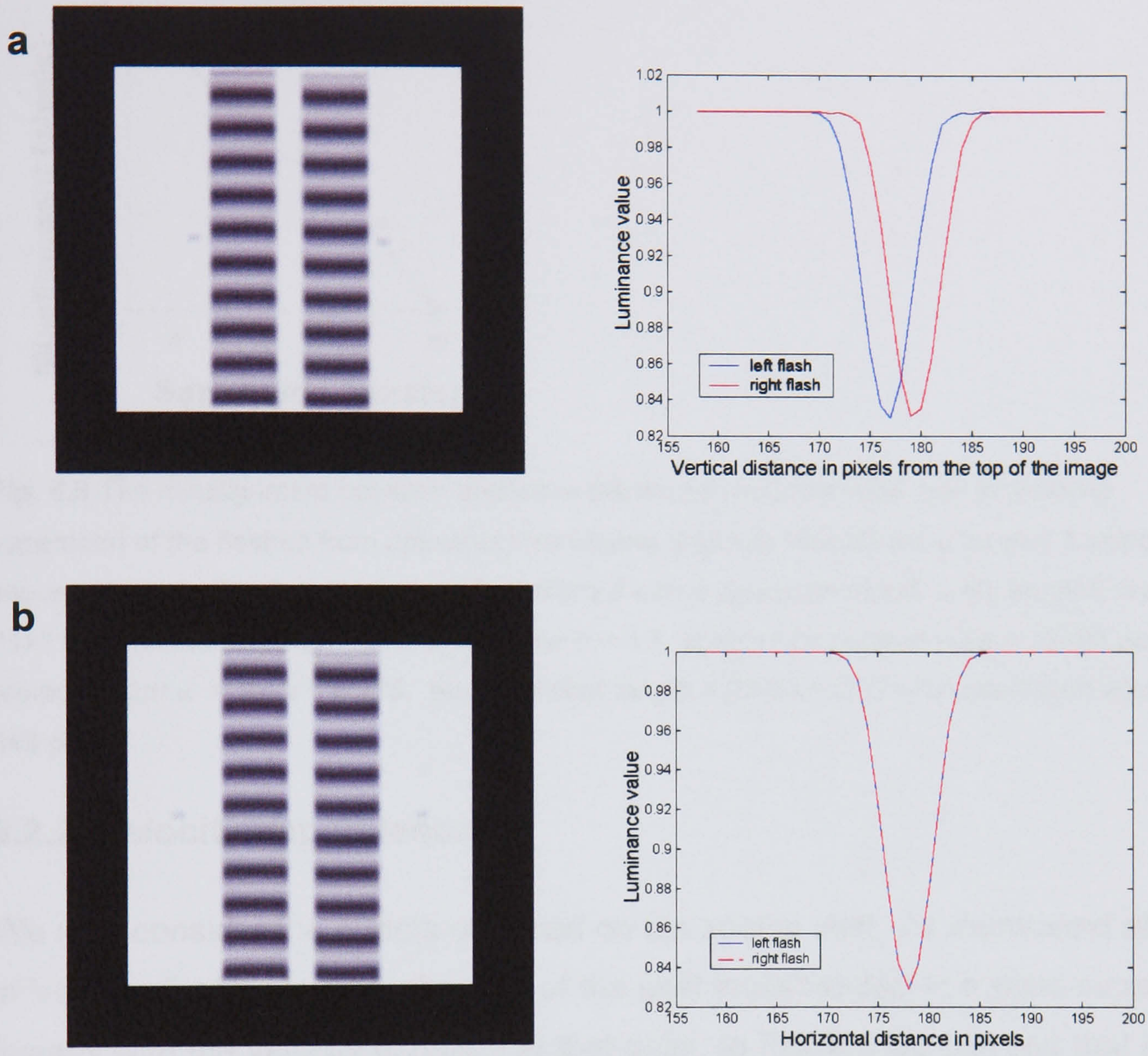


Fig. 6.4 Presenting translating gratings (left – upwards, right – downwards) with flashes either side, to Version 5 of the model. Motion quotient blurred with a Gaussian of s.d. = 40, support area = 71×71 pixels. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Reconstruction of the frame in which the flash representation peaks and a plot of the luminance along the vertical line through the middle of the flashes. Separation of flashes from gratings: 8 pixels, misalignment: 2.6 pixels. (b) Separation: 29 pixels, misalignment: 0 pixels.

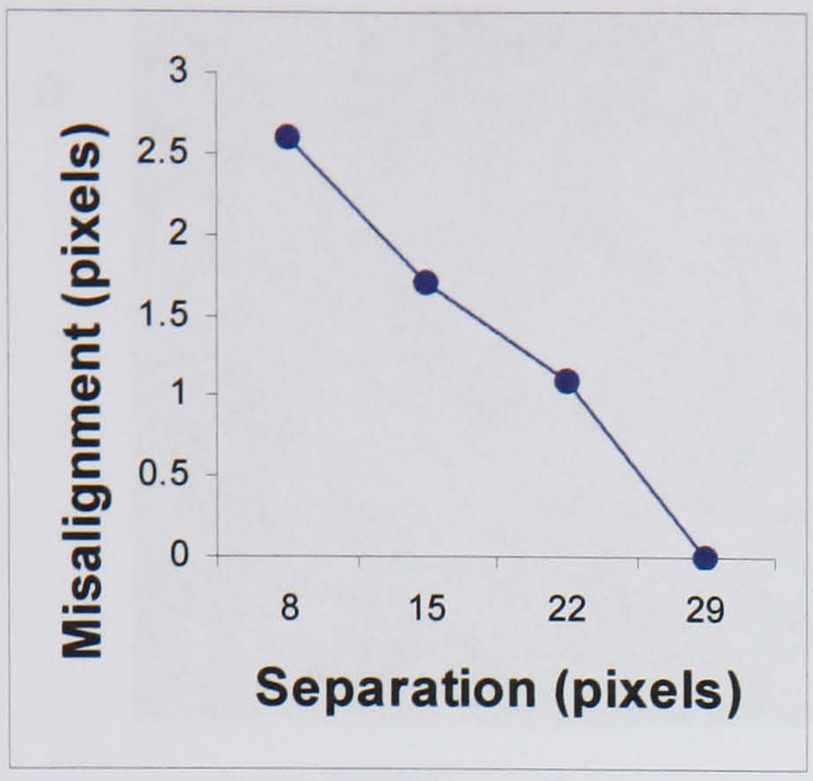


Fig. 6.5 The misalignment between flashes in the model reconstruction over increasing separation of the flashes from oppositely translating gratings. Results from version 5 using sequences as in Fig. 6.4. Motion quotient blurred with a Gaussian of s.d. = 40, support area = 71×71 pixels. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels.

6.2.2 Velocity dependence

We now consider the effects of speed on the spatial shift. As mentioned above, in Version 5 of the model, the size of the shift implemented at a point increases linearly with the velocity estimate at that point. In Fig. 6.6 we can see that in the case of Gabor patches drifting in opposite directions at a velocity of 3 pixels/frame, the resulting misalignment is 6 pixels, i.e. 1.5 times what we found with a drifting velocity of 2 pixels/frame. (The Gabor patterns are more blurred than in previous examples as the higher velocity causes more blurring in time.) However, in the original experiment (De Valois & De Valois, 1991) although there was at first some increase with speed, this effect levelled off at higher speeds, rather than increasing linearly.

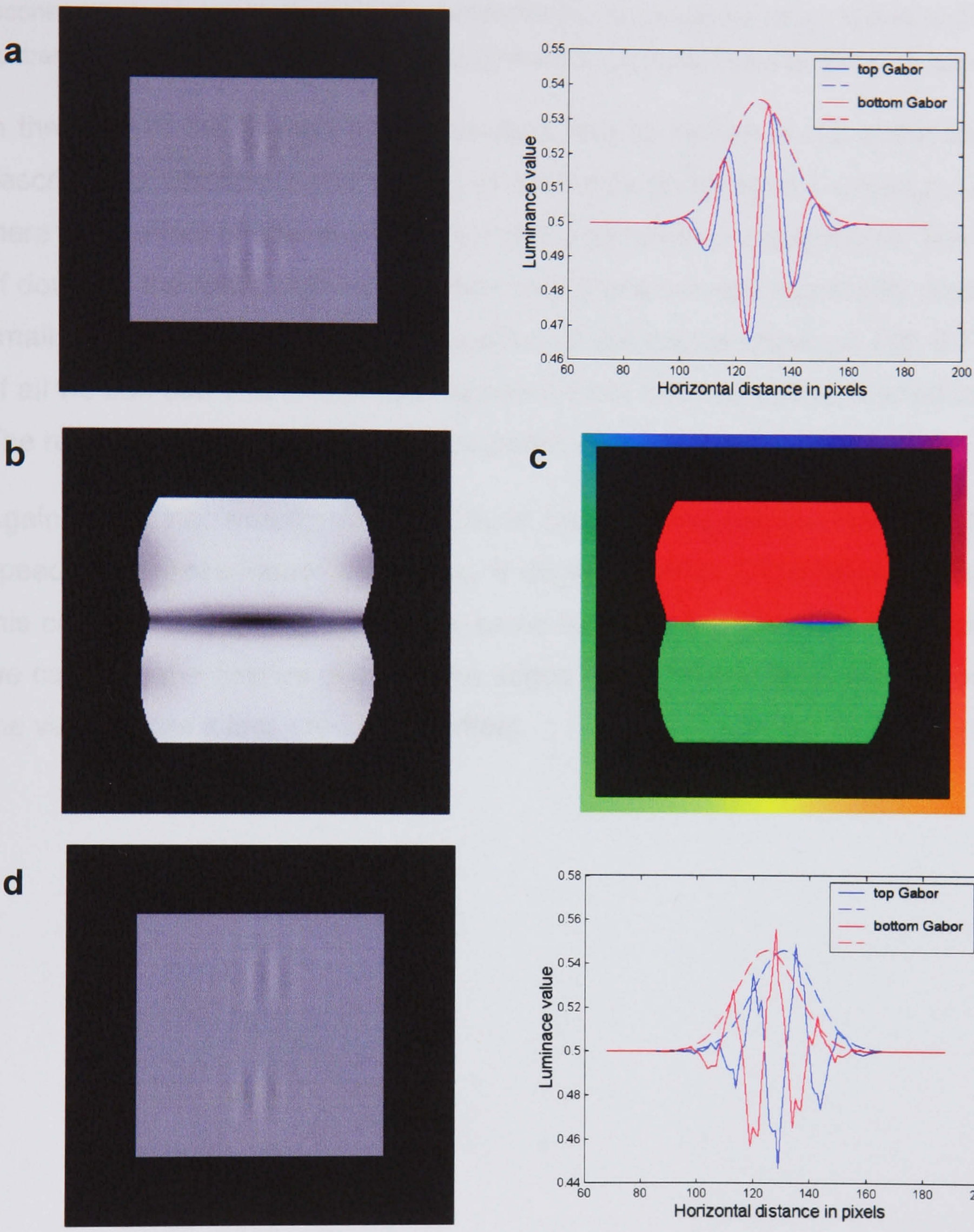


Fig. 6.6 Drifting Gabor patches presented to version 5 of the model (top – rightwards, bottom – leftwards). Drifting velocity of 3 pixels/frame (1.5 times the speed of previous examples). Gaussian velocity quotient blur of s.d. = 40, support area = 71×71 pixels. Model parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels; temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) Single frame from blurred output shown along with the plot of a horizontal line of luminance values drawn through the centres of each of the Gabor patches. (b) Velocity magnitude output, (4 pixels/frame – white, 0 pixels/frame – black). (c) Velocity direction output. (d) Corresponding frame from

reconstructed output with the plot of a horizontal line of luminance values drawn through the centres of each of the Gabor patches. Misalignment of 6.0 pixels found by fitting Gaussians.

In the case of the flashed bars presented near to motion in the empirical work described in Chapter 2 and in past experiments (Whitney & Cavanagh, 2000), there is no effect of speed on the size of the perceived misalignment. The effect of doubling the (anticlockwise) rotation speed of a bar with two briefly presented small bars either side (at the 60° position of the bar) is shown in Fig. 6.7. First of all we can see that with greater speed the bar becomes more blurred in time. The reconstruction still preserves the shape of the blurred bar.

Again, a greater misalignment is found between the two flashes at a higher speed (3 degrees/frame: 1.1 pixels, 6 degrees/frame: 1.6 pixels), although in this case the increase is not by the same factor as the increase in velocity. As we can see, the flashes occur at the edges of the motion field, where doubling the velocity has a less predictable effect.

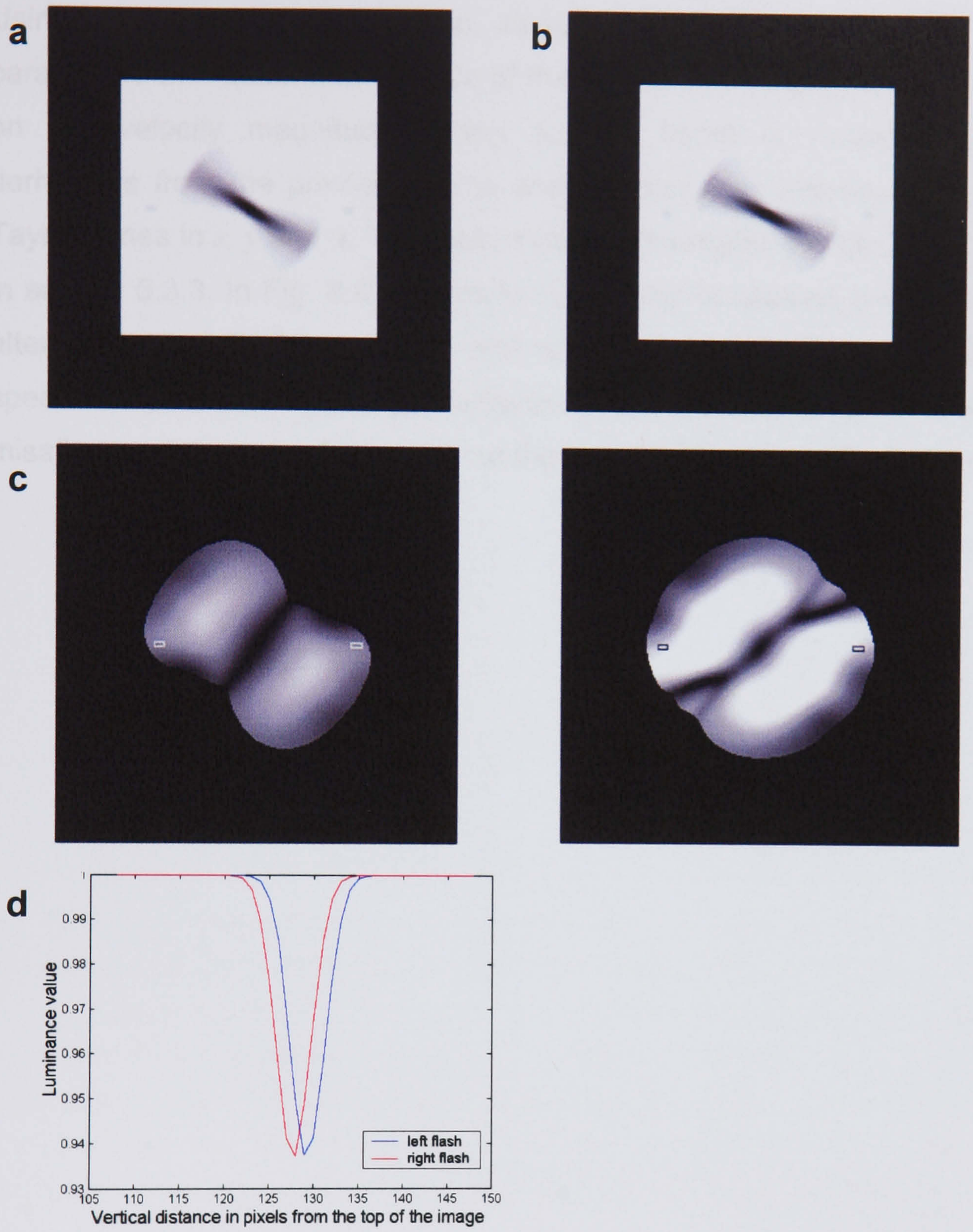


Fig. 6.7 A bar rotating anticlockwise at 6°/frame, results from version 5. Gaussian velocity quotient blur of s.d.= 40 and 71×71 pixel support area is applied. Model parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels, temporal filter parameters: $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Taylor expansion window = 3×3 pixels. (a) The blurred image from the frame with the peak flash response. (b) The corresponding reconstructed image. (c) The velocity for a rotation of 3°/frame (left) is compared to the velocity output for 6°/frame (right). (d) The plot shows values along each of the vertical columns running through the middle of each of the flashes from the top of the image to the bottom. There is a misalignment consistent with the directions of motion of 1.6 pixels.

Using the temporal reconstruction, as suggested in Section 4.5, the feedback parameters that determine the size of the shift are no longer linearly dependent on the velocity magnitude. Every second frame is reconstructed using derivatives from the previous frame and the motion is incorporated in the 3D Taylor series in x, y and t. The motion feedback weights are given by Eqns. 5.3 in section 5.3.3. In Fig. 6.8 the results for drifting Gaussians are shown for the altered Version 5 of the model, with temporal reconstruction, comparing the speed of 2 pixels/frame to 3 pixels/frame. There is a much smaller difference in misalignment than was found without the temporal reconstruction alteration.

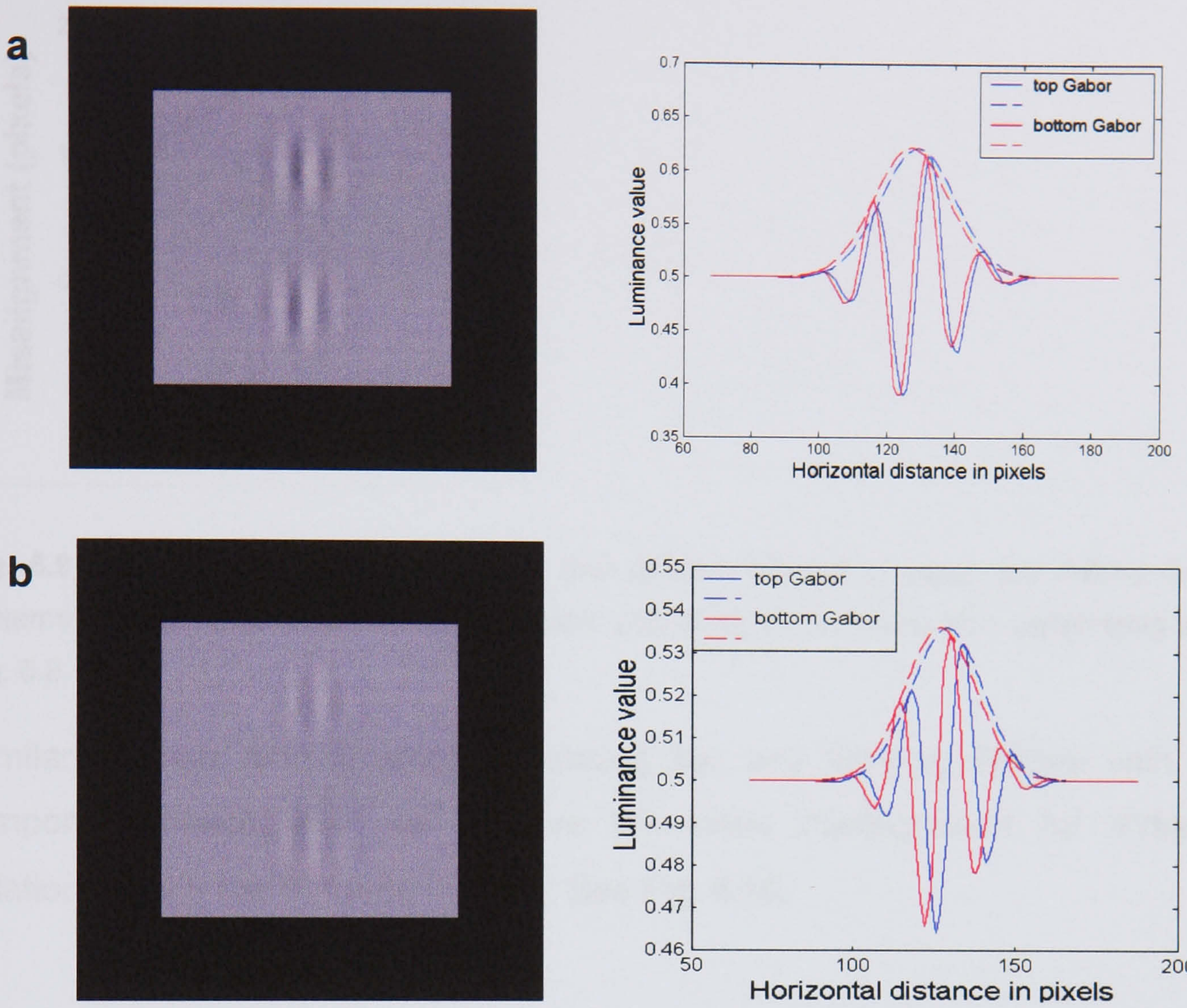


Fig. 6.8 Drifting Gabor patches (top - rightwards, bottom - leftwards) as input in Version 5, with temporal reconstruction introduced, so that every second frame is reconstructed. Gaussian velocity quotient blur of s.d. = 40, support area = 71×71 pixels is applied. Model parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels; temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. (a) Velocity = 2 pixels/frame. The reconstructed frame is shown alongside the plot of the luminance horizontally through the middle of each flash. Misalignment found by fitting Gabors was 2.1 pixels. (b) Velocity = 3 pixels/frame, reconstructed frame shown alongside the plot of the luminance horizontally through the middle of each flash. Misalignment found by fitting Gabors was 2.2 pixels.

The results for four different speeds of drifting Gabor patterns are plotted below (Fig. 6.9) and we see the same pattern as was found experimentally (De Valois & De Valois, 1991). There is an initial increase in misalignment that levels off with higher speeds.

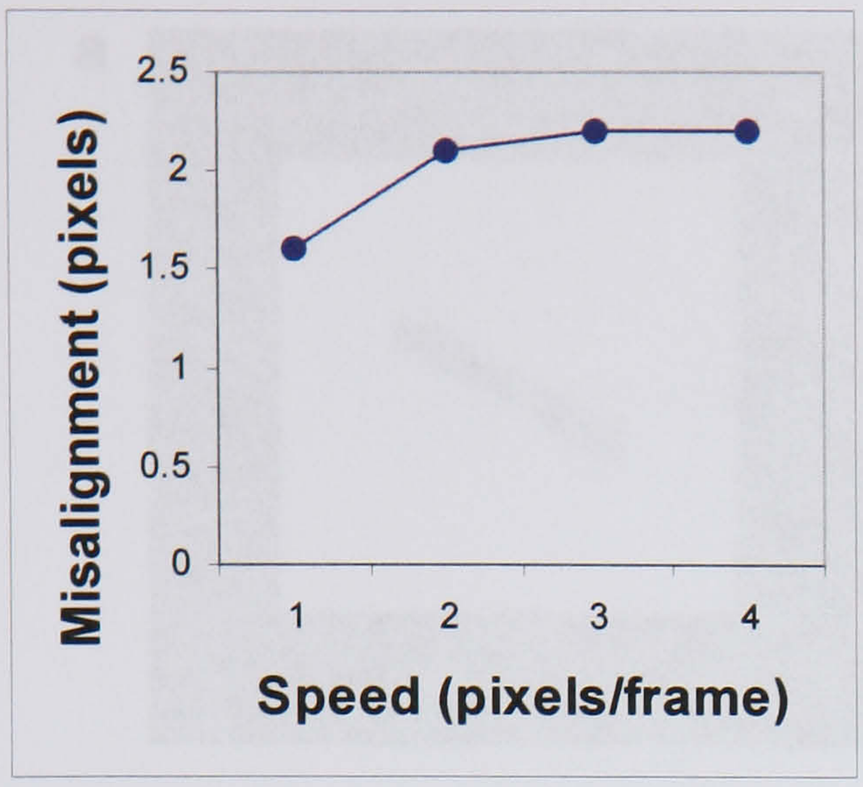


Fig. 6.9 Relative misalignment results for four different speeds of oppositely drifting Gabor patterns as found from Version 5 of the model, with temporal reconstruction, parameters as in Fig. 6.8.

Similarly, when reconstructing a rotating bar and flanking flashes with the temporal reconstruction we observe the same misalignment for 3°/frame rotation as with the 6°/frame rotation. See Fig. 6.10.

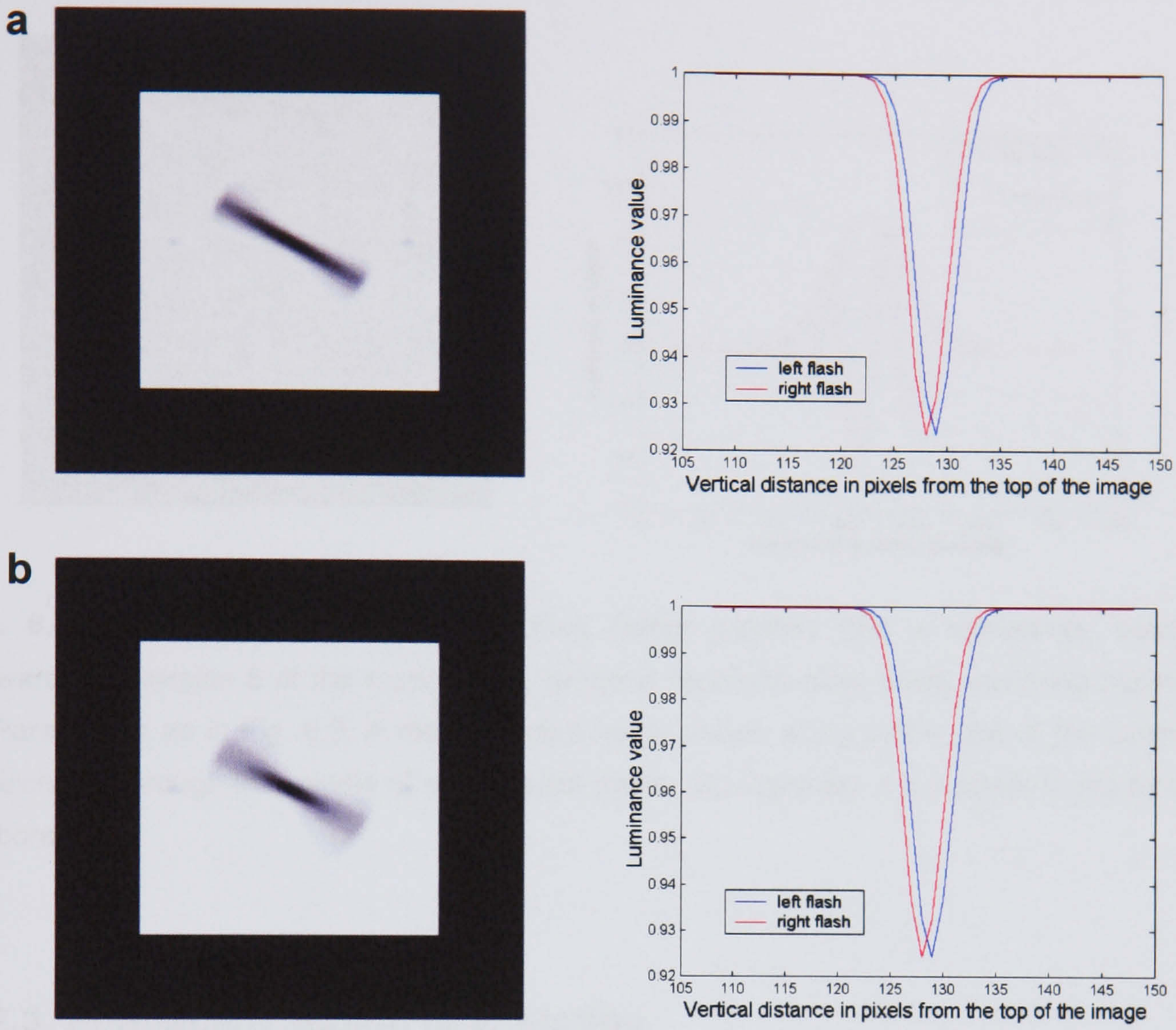


Fig. 6.10 Anticlockwise rotating bar with flashed bars presented either side at the 60° position of the bar, presented to Version 5 of the model. Using temporal reconstruction, so that every second frame is reconstructed. A Gaussian velocity quotient blur of s.d.= 40, support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. (a) Bar rotating at 3°/frame. Reconstruction for the frame in which the flash representation peaks shown alongside the luminance plotted along the vertical line through the flashes. Misalignment of 0.8 pixels found by fitting Gabors. (b) Bar rotating at 6°/frame. Misalignment = 0.8 pixels.

The induced shift is much smaller than the results from the previous model. The shift can be made comparable in size by increasing the constant that relates the three weights in Eqns. 5.3 in section 5.3.3. The results for the weighting constant = 2 is shown in Fig. 6.11 for the two drifting Gabor patches. A misalignment of 2.9 pixels is found.

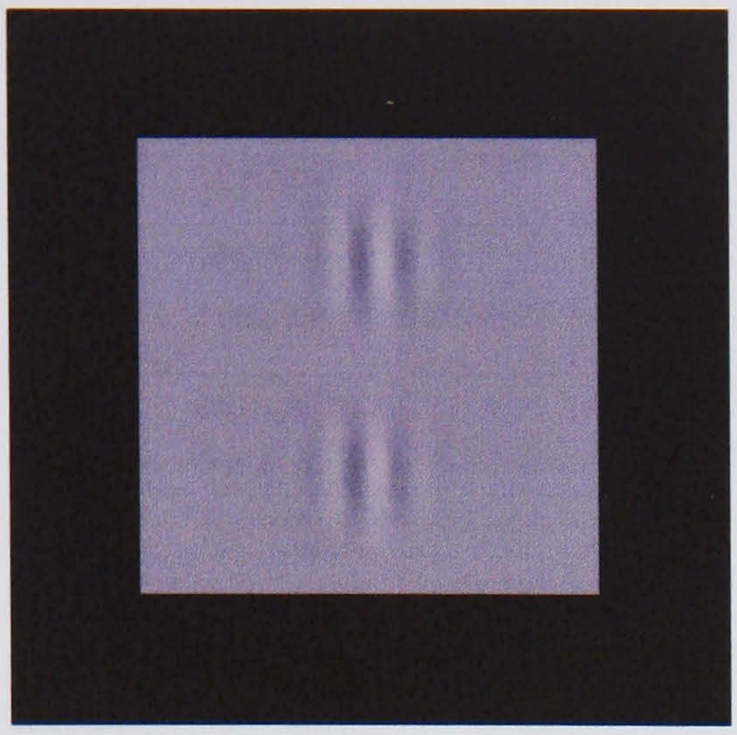


Fig. 6.11 Presenting two oppositely drifting Gabor patches (top – rightwards, bottom – leftwards) to version 5 of the model. With temporal reconstruction, using weighting constant = 2. Parameters as in Fig. 6.9. A reconstructed frame shown along with a plot of the luminance horizontally through the middle of each Gabor patch. Misalignment = 2.9 pixels found by fitting Gabors.

6.2.3 Permanent stimuli near motion

The next step was to examine the output of the model if small flanking bars were present permanently next to motion. This experimental condition does not result in a perceived misalignment for subjects (Whitney & Cavanagh, 2000). However, in this case the model fails to predict the experimental results. This is shown for the case of oppositely drifting sine gratings with flanking bars in Fig. 6.12. The small bars are less misaligned then when presented for only one frame. This is because a briefly presented bar has an effect of increasing the motion values around it, which is not the case for a permanent bar. However, the static bars are still shifted in the reconstruction. A misalignment also occurs between permanent flankers in the version with temporal reconstruction. See Fig. 6.13.

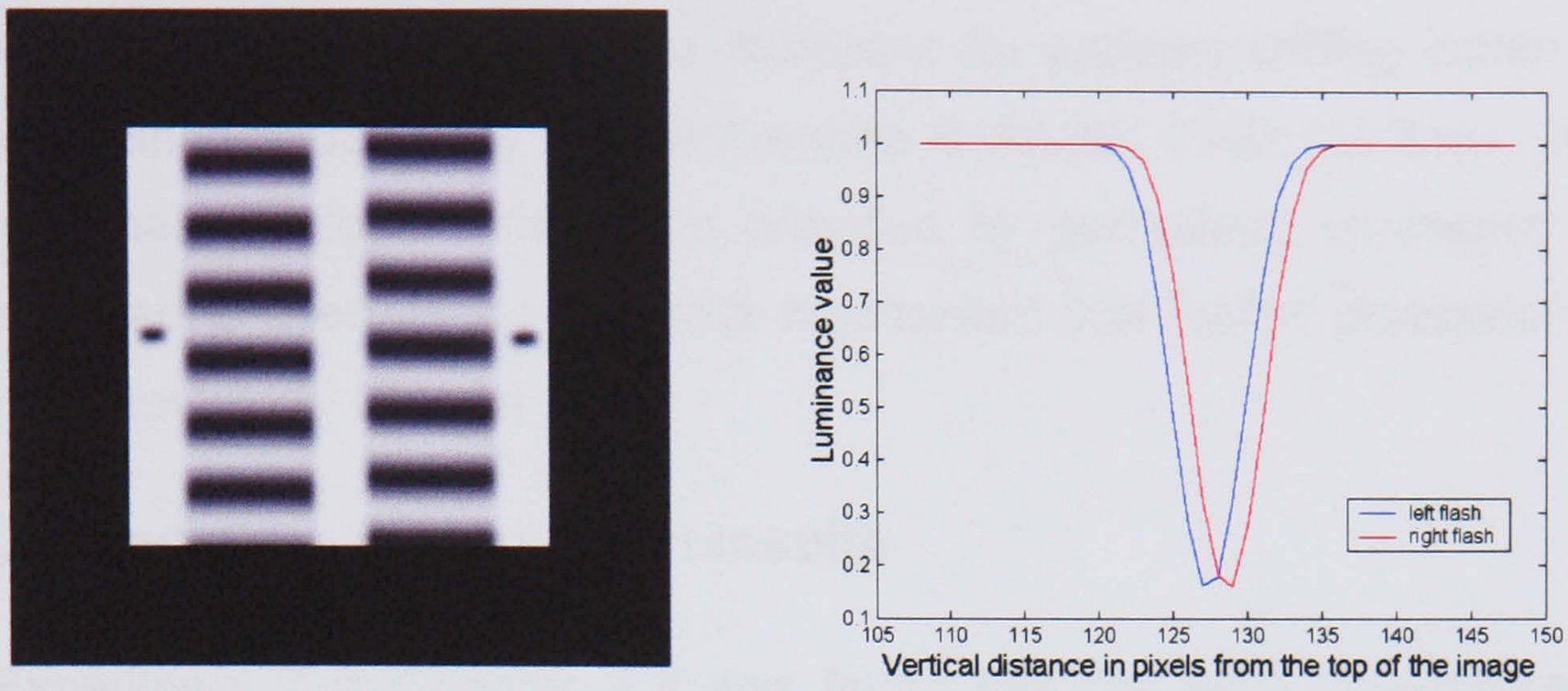


Fig. 6.12 One frame of reconstructed output from Version 5 of the model for translating gratings with permanent small bars presented either side. Velocity quotient blurring of s.d.=40 , support area =71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. Shown on the right is a plot of the vertical column of luminance values through the middle of each of the two flashes. A misalignment of 1.2 pixels was found. A 2 pixel misalignment was found with temporary flashes.

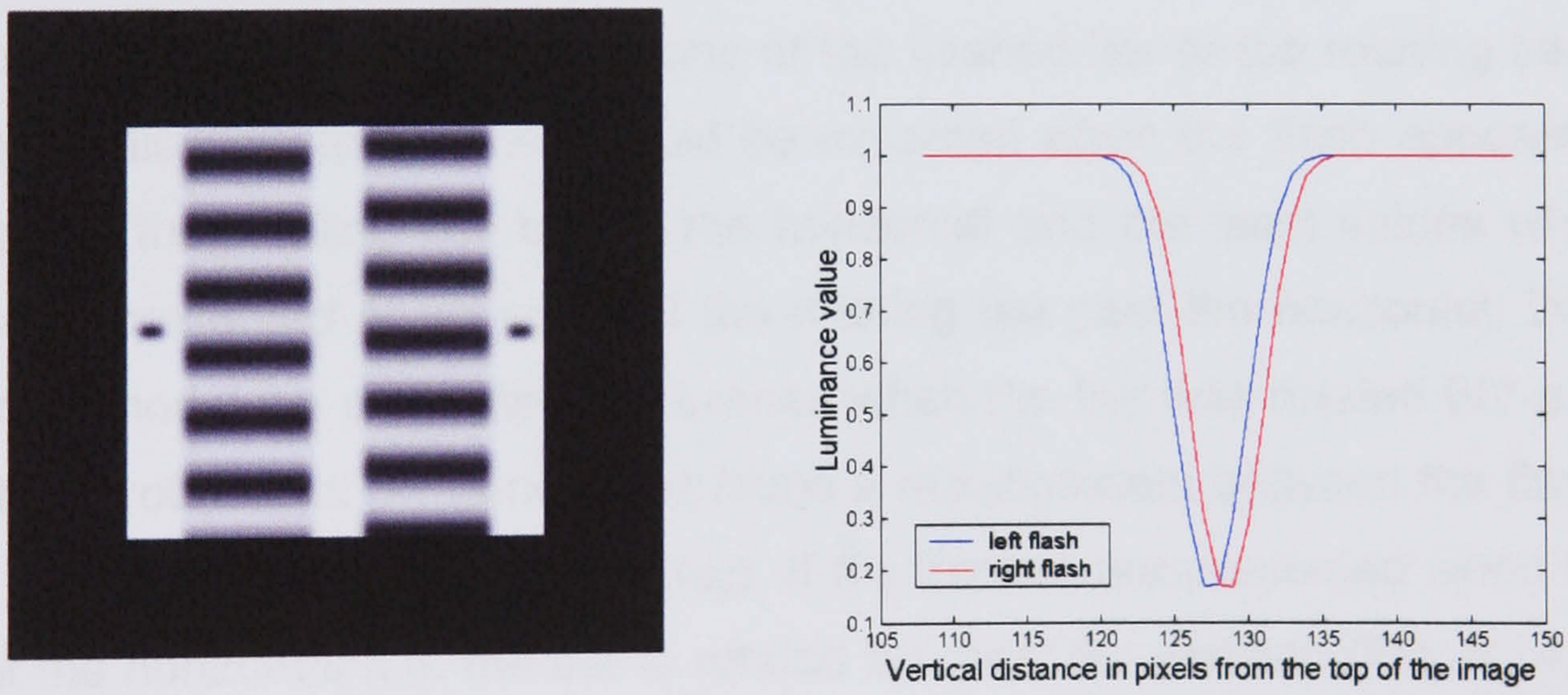


Fig. 6.13 One frame of reconstructed output from Version 5 with temporal reconstruction for translating gratings with permanent small bars presented either side. Velocity quotient blurring of s.d.=40, support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. On the right is the vertical column of pixel values through the middle of each of the two flashes. A misalignment of 1.1 pixels was found.

The shift has also been found to disappear for patterns drifting within a hard edge luminance boundary (Ramachandran & Anstis, 1990; Whitney, 2003). It seems that additional information provided by permanent luminance values would need to override the feedback mechanism. For further discussion of this see Further Work, Chapter 8.

6.2.4 Modelling the empirical results

In Experiment 2 in Chapter 2 it was found that the flash misalignment was greatest if the flashes were presented before the rotating bar reached the position of the flashes. The size of the effect of the bar on the perceived position of the flash was not symmetric around the horizontal. It was suggested this was due to the priming of large motion cells by motion at a distance, which then influenced the spatial representation of the flash as it was becoming established. We now examine this in the context of the motion feedback model presented above. From the experimental results we would expect misalignment to occur at several nearby positions of the flashed bar to the rotating bar. The largest misalignment values would be expected when the flash appears at an angle of the rotating bar before the horizontal and the least values when the flash is presented at an angle of the rotating bar past the horizontal. With the current model we presented the flashes when the bar was rotated 60° past the vertical (rotating at 3°/frame), and found a misalignment between the flashes in the representation (see Fig. 6.10(a)). If the flashes are presented when the bar is at the horizontal (i.e. the bar is rotated 90° past the vertical) (Fig. 6.14), there is a larger misalignment than for 60°. For a sequence where the flashes are presented when the bar is rotated 150° past the vertical (Fig 6.15), there is less misalignment between the flashes than for 60° and 90°.

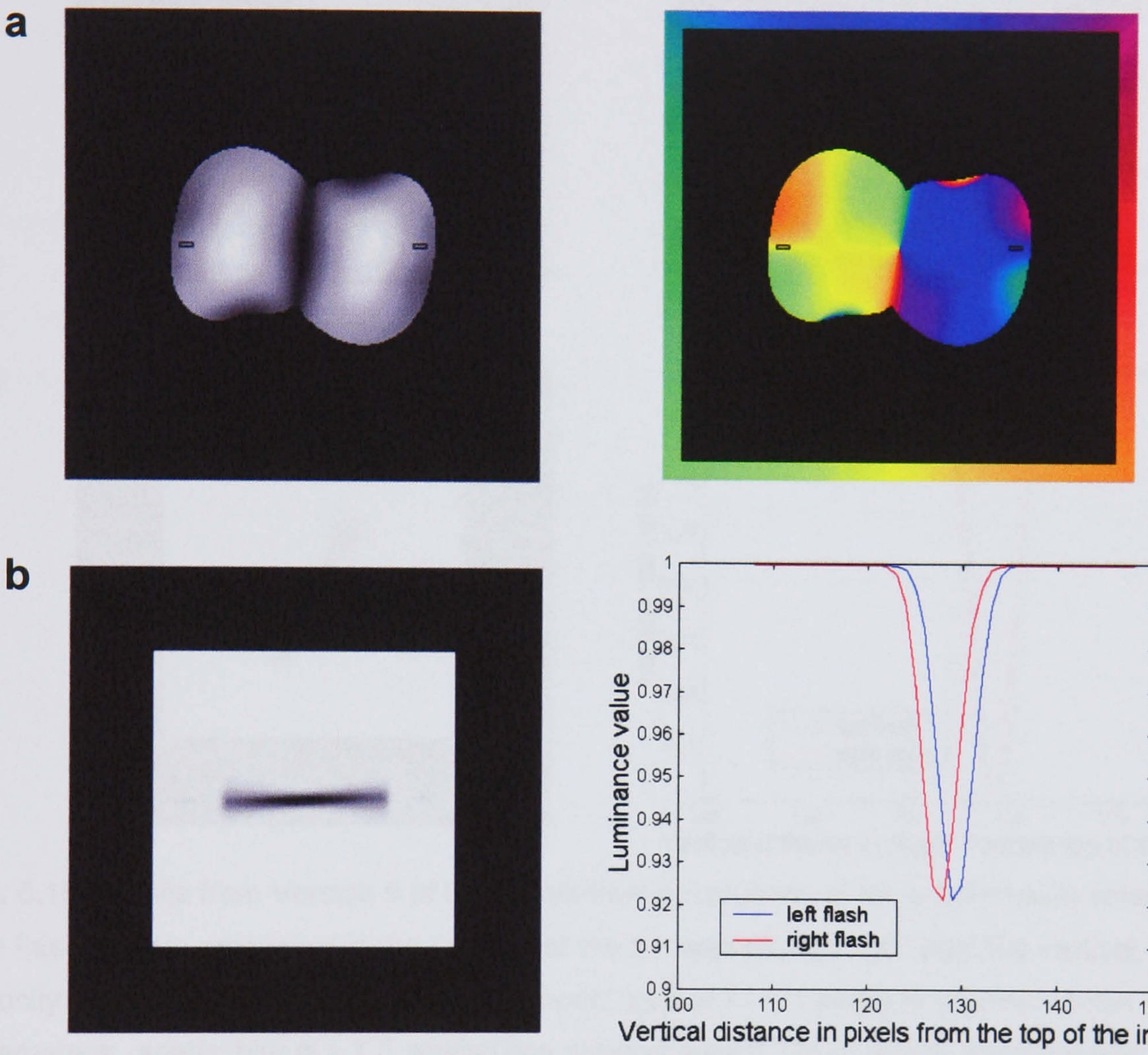


Fig. 6.14 Results from Version 5 of the model, for a sequence of an anticlockwise rotating bar. The flashes are present in the frame that the bar is horizontal. Velocity quotient blurring of s.d.=40 and support area=71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels, temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max: 1.6 pixels/frame - white, 0-black) and direction. (b) Reconstruction of the frame in which the flash representations peak. On the right is shown the plot of the vertical column of luminance values through the middle of each of the flashes. Relative misalignment of 1.6 pixels was found by fitting Gaussians.

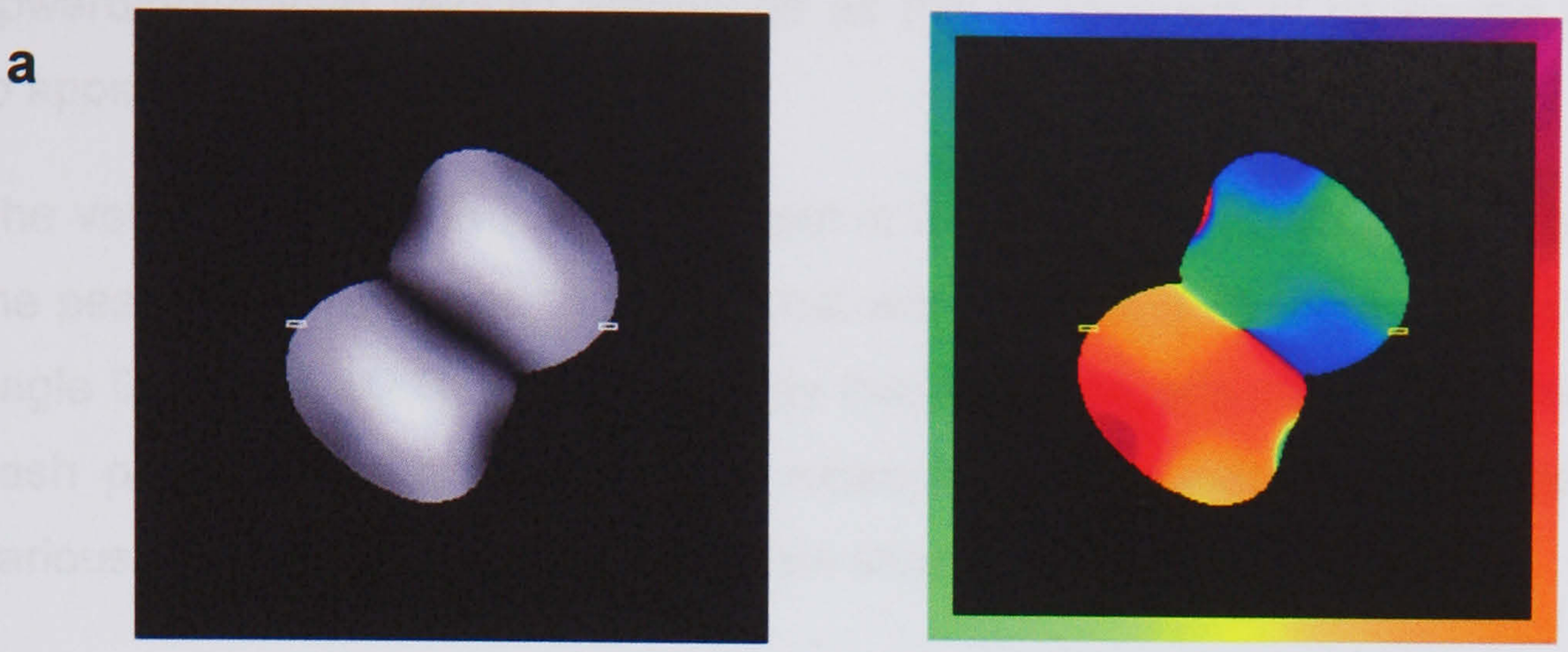


Fig. 6.15 Results from Version 5 of the model from a sequence of an anticlockwise rotating bar. The flashes were presented in the frame that the bar was rotated 150° past the vertical. Velocity quotient blurring of s.d.=40 and support area of 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. (a) Velocity magnitude (max 1.6 pixels/frame - white, 0 - black) and direction. (b) Reconstruction of the frame in which the flash responses peak. On the right the plot of the vertical column of luminance values through the middle of each of the flashes is shown. Relative misalignment of 0.5 pixels found by fitting Gaussians.

In order to understand the pattern of motion influence on spatial position over different locations of the rotating bar, we will initially consider the motion output from a moving bar by itself. In Fig. 6.16 the velocity output over different angles of rotation of the bar – without the presentation of the flashes – is shown for a speed of rotation of 3°/frame. In this case we are interested specifically in the

upward/downward velocity magnitude as this is what would cause the flashes to appear misaligned.

The velocity at a given frame of output is the velocity that would be present at the peak of the response to a flash that was presented 10 frames earlier, at an angle θ of the bar. This is the velocity that would be affecting the position of a flash presented at angle θ. A few examples of the velocity associated with various angles of flash presentation are shown in Fig. 6.16.

The velocity outputs show that as the bar rotates round, the velocity field overlaps the area of the flashes for part of the trajectory. Note that the velocity field is not symmetric around the bar as one might originally expect. In the motion field corresponding to θ = 90°, for example, we see a longer trailing end to the motion around the bar and an earlier peak. This is due to the shape of the temporal filters, which are skewed in time.

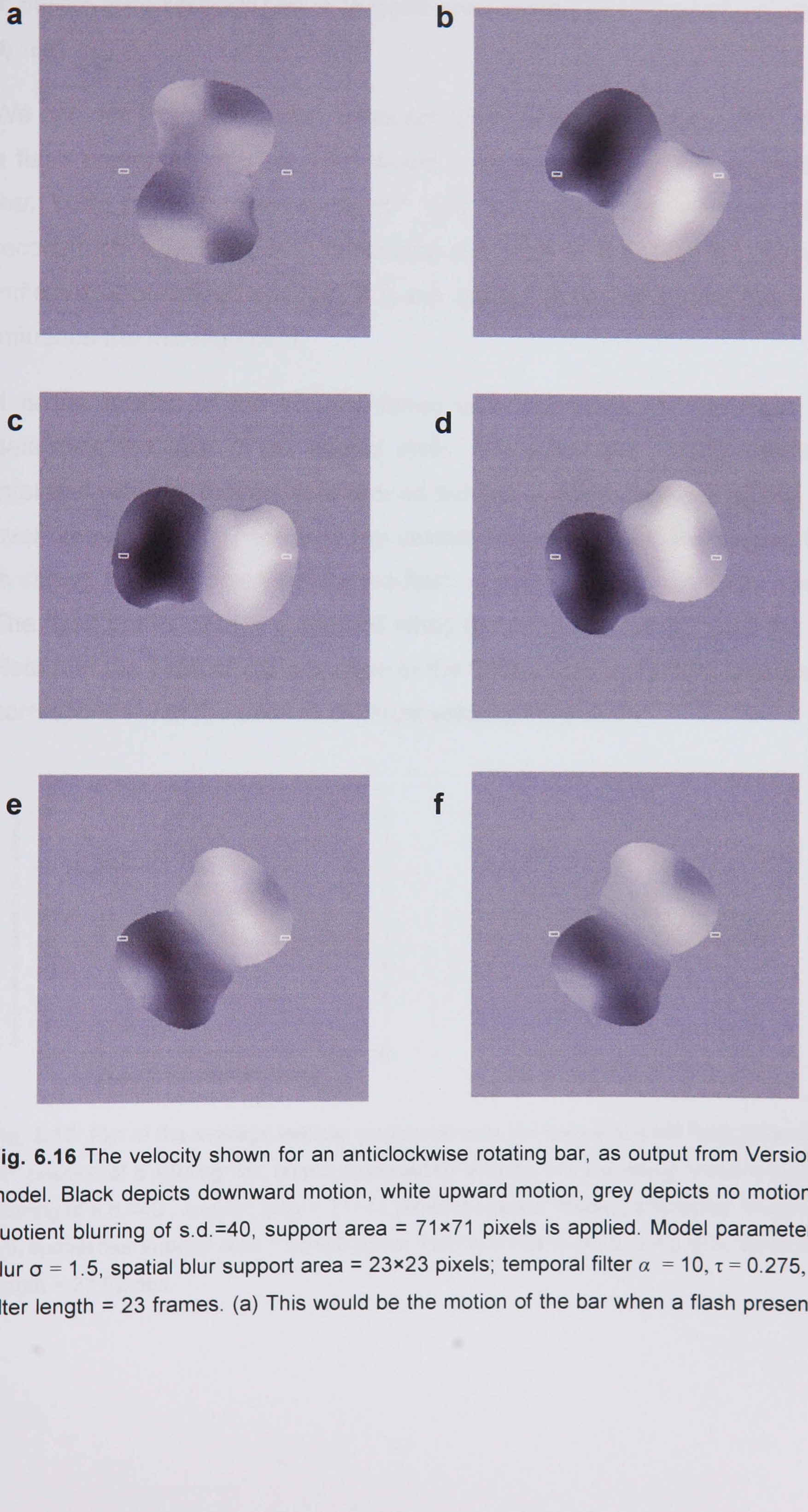


Fig. 6.16 The velocity shown for an anticlockwise rotating bar, as output from Version 5 of the model. Black depicts downward motion, white upward motion, grey depicts no motion. Velocity quotient blurring of s.d.=40, support area = 71×71 pixels is applied. Model parameters: spatial blur $\sigma = 1.5$, spatial blur support area = 23×23 pixels; temporal filter $\alpha = 10$, $\tau = 0.275$, temporal filter length = 23 frames. (a) This would be the motion of the bar when a flash presented at the

0° position of the bar would peak in its spatial representation. (b) 60°, (c) 90°, (d) 120° (e) 144° (f) 156°.

We can see from the motion areas produced at different angles (Fig. 6.16) that a flash presented at 0° or 156° would have no bar velocity overlapping it and that both those presented at 60° and 90° would. As we are taking the reconstruction of the frame containing the peak of the flash response as our indicator of perceived position, it is the motion of the bar in this frame that will influence the misalignment.

It is the motion of the bar combined with the motion of the flash that will determine the size of the spatial shift. We now need to see how the flash interacts with the motion area around the bar – which changes spatial location over time. In Fig. 6.17 the average velocity magnitude over the area of the flash is shown from 5 frames before the flash is presented to 30 frames afterwards. The flash in this case is presented when the bar is rotated 60° past the vertical. Note that the peak of the response to the flash in frame 10 after its presentation corresponds with the peak in the local velocity magnitude.

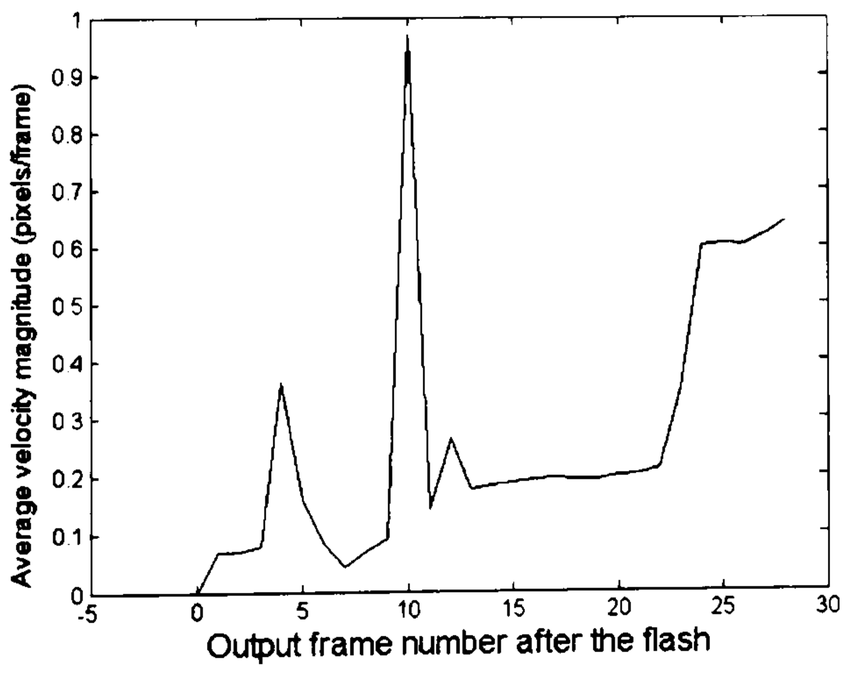


Fig. 6.17 Plot of the average velocity magnitude over the area of the left flash presented at the 60° position of a rotating bar, output produced by Version 5 of the model. Velocity quotient blurring of s.d.=40 , support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames.

Presented below are the misalignment magnitudes at various angles of presentation of the flash - as predicted by the model. There is no misalignment between angles 156° - 48° (Fig. 6.18). This reflects the spatial overlap of the bar's motion with the position of the flashes at the different angles. The misalignment magnitude values are centred roughly around the horizontal, although they are not symmetrical around 90°. As is predicted by the motion field in Fig. 6.16 (e), there is an increase in the size of the misalignment between the flashes that are presented after the horizontal position of the rotating bar, as shown by the jutting lower arm of the data in Fig. 6.18. This is not found in the experimental results in Chapter 2. Plotting the results on polar coordinates makes for an easier comparison. There is a similar pattern as the results of Experiment 1 in Chapter 2, but the misalignment values are centred around the horizontal and for flashes presented after the horizontal position of the bar there is an increase in misalignment in the model results, rather than just a decrease after the peak as in the experimental results.

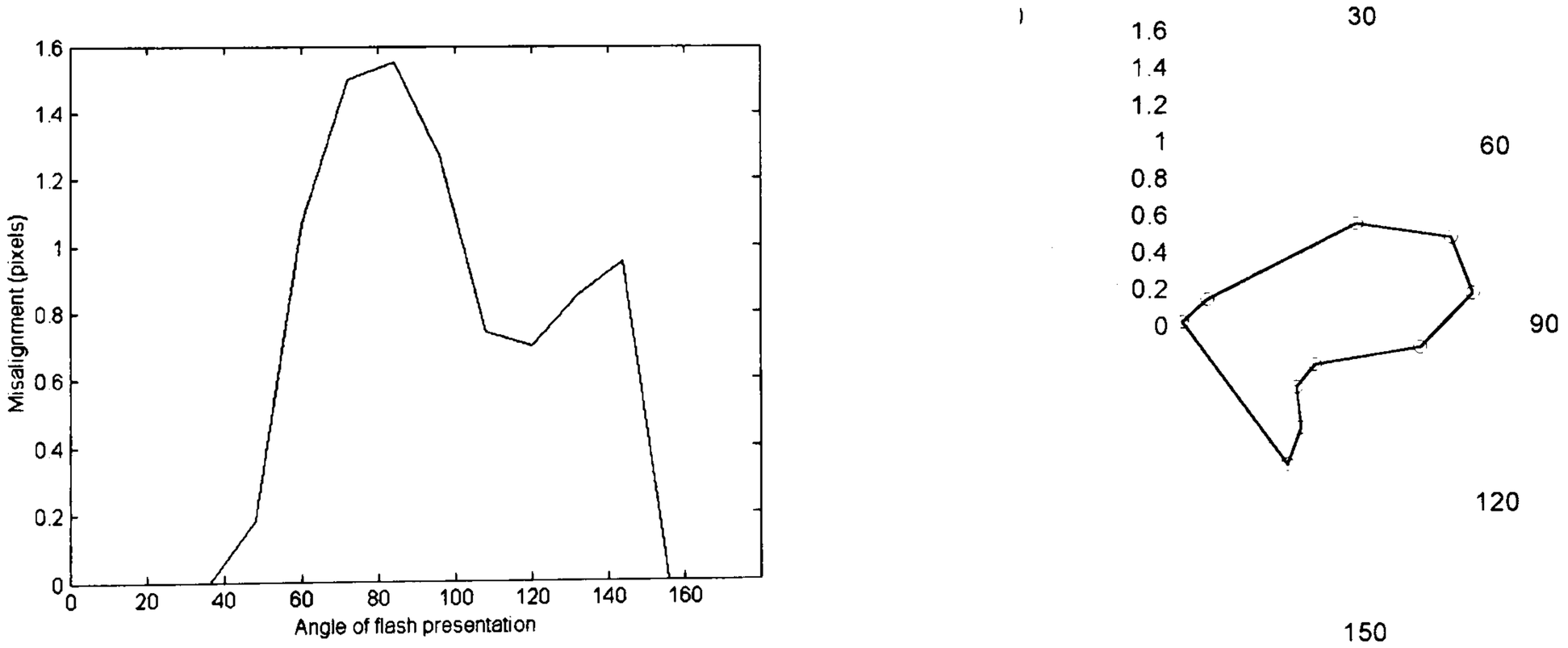


Fig. 6.18 The misalignment between flashes flanking a rotating bar (3°/frame) plotted over angles of presentation, as found by the model Version 5. Results from several sequences of an anticlockwise rotating bar. Velocity quotient blurring of s.d. = 40, support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels.

We now consider how this pattern might change with a faster speed of rotation. With a faster speed of rotation (see Fig. 6.19(a) for 4°/frame), the extent of the motion field (in the direction of motion of the bar) is greater. The motion field takes the same shape, so that in effect the same pattern of results is spread out over a wider range of angles. In the experimental data, the misalignment values do become less sharply tuned over different angles in the same way.

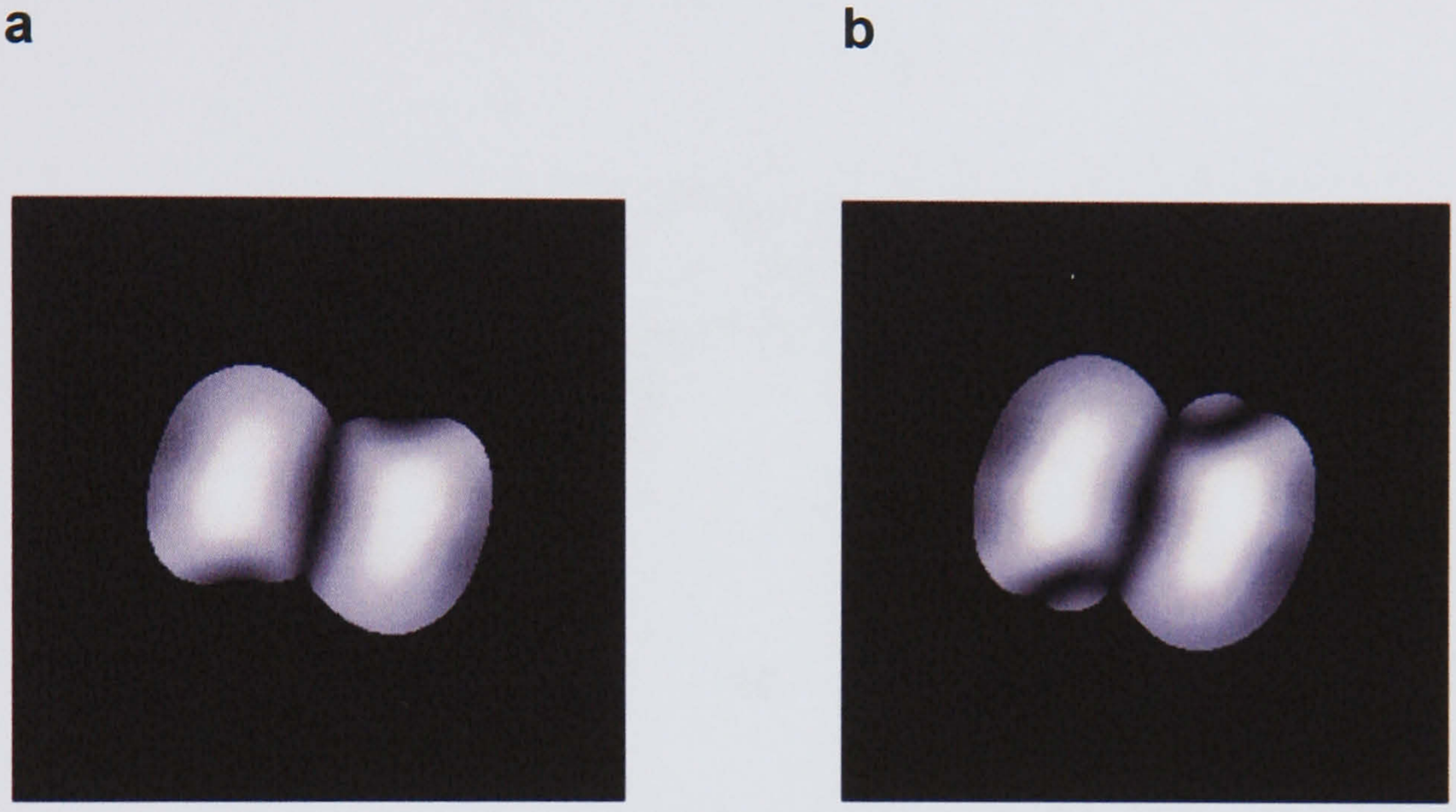


Fig. 6.19 (a) The velocity magnitude estimate of a bar rotating anticlockwise at 3°/frame. Max =1.6 – white, min = 0 – black. (b) The velocity magnitude estimate of a bar rotating anticlockwise at 4°/frame. Max =2.1 – white, min = 0 – black. Results from version 5 with parameters as in Fig. 6.18.

The larger spatial extent of motion influence at higher speeds is a result of the fact that the temporal filters contribute to the spatial extent of the motion field. The length of the temporal filter remains the same for all speeds, so at higher speeds a longer spatial trajectory of the rotating bar is convolved over the same time.

Below we plot the resulting misalignments from three different speeds of rotation. The original Version 5 of the model, using only spatial reconstruction, does not reproduce the rotation of the peak effect away from the horizontal with

increased speed that was observed experimentally. It also results in an increased misalignment over different speeds as we also saw earlier.

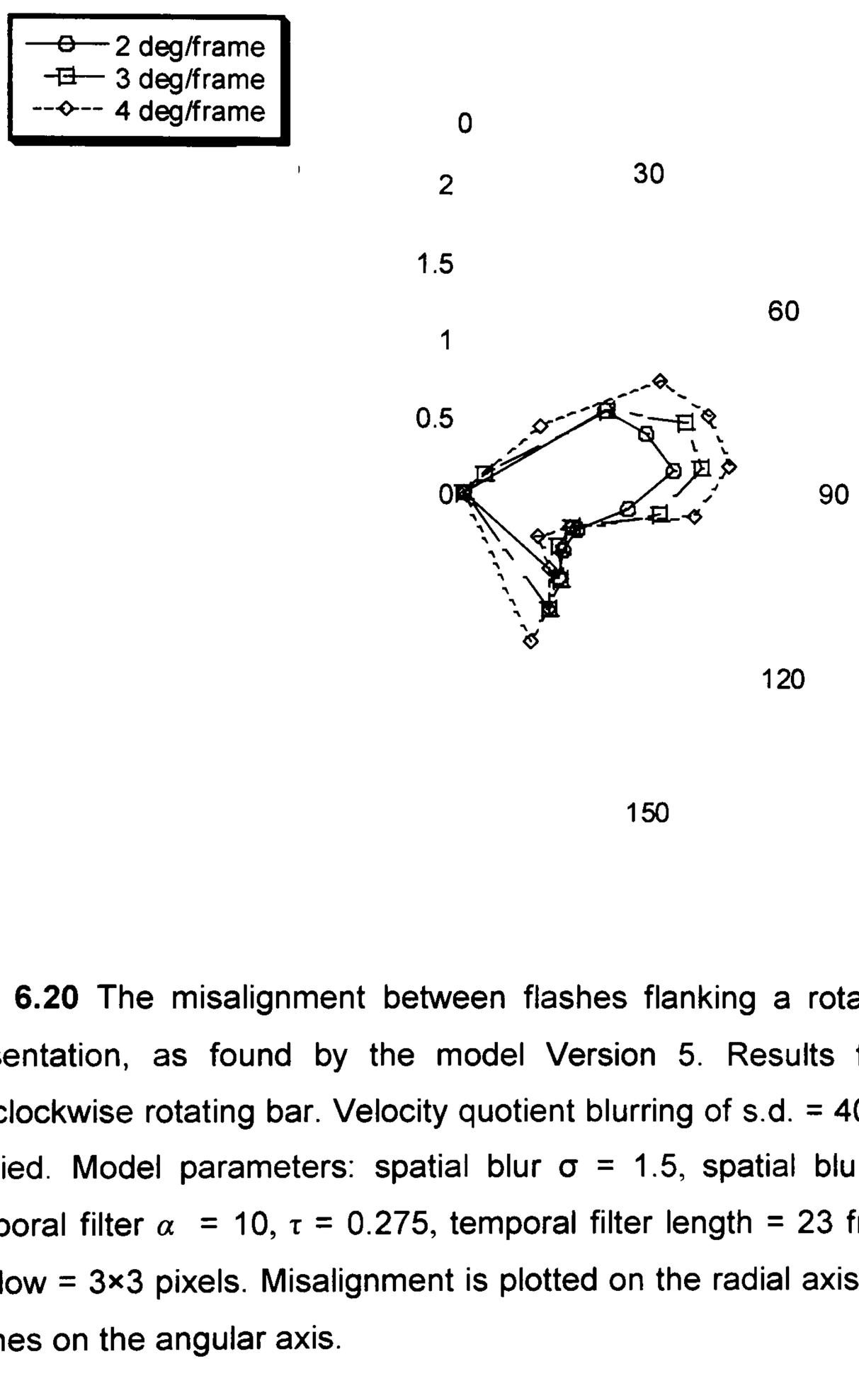


Fig. 6.20 The misalignment between flashes flanking a rotating bar plotted over angles of presentation, as found by the model Version 5. Results from several sequences of an anticlockwise rotating bar. Velocity quotient blurring of s.d. = 40, support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. Misalignment is plotted on the radial axis against angle of presentation of flashes on the angular axis.

If we use the model with the temporal reconstruction, so that the weights are corrected for as mentioned in Section 5.3.3 and every second frame is completely reconstructed, using weights in time in the Taylor reconstruction, then the higher velocity no longer leads to a greater misalignment as seen in Section 6.2.2. The results for this version of the model are shown below in Fig. 6.21. We can see that although there is no increase in the size of the effect

over velocities, as was observed experimentally, there still appears to be no effect of speed on the peak misalignment.

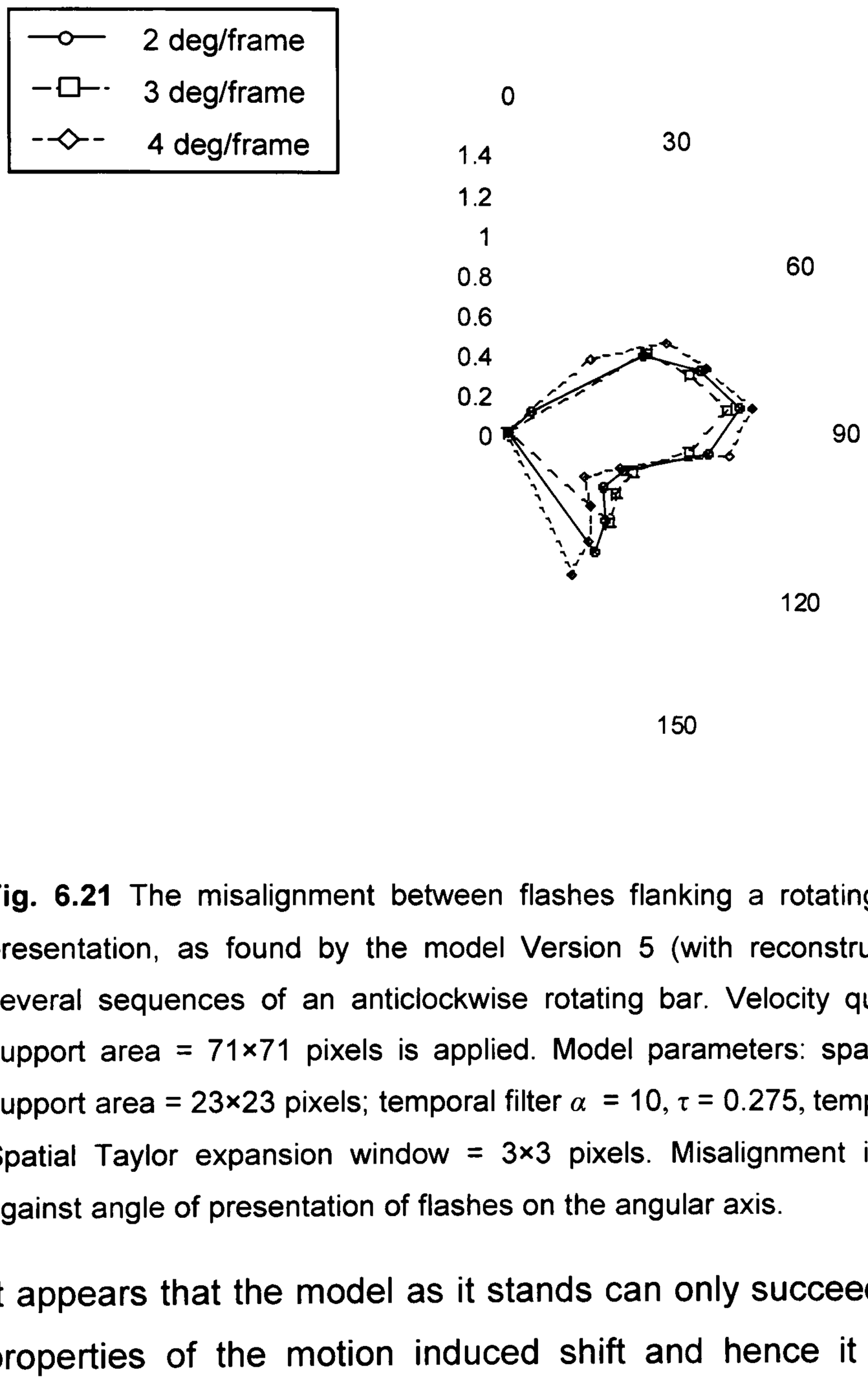


Fig. 6.21 The misalignment between flashes flanking a rotating bar plotted over angles of presentation, as found by the model Version 5 (with reconstruction in time). Results from several sequences of an anticlockwise rotating bar. Velocity quotient blurring of s.d. = 40, support area = 71×71 pixels is applied. Model parameters: spatial blur σ = 1.5, spatial blur support area = 23×23 pixels; temporal filter α = 10, τ = 0.275, temporal filter length = 23 frames. Spatial Taylor expansion window = 3×3 pixels. Misalignment is plotted on the radial axis against angle of presentation of flashes on the angular axis.

It appears that the model as it stands can only succeed in capturing the spatial properties of the motion induced shift and hence it reproduces the general shape of the effect of a rotating bar on a briefly presented flash at different points along its trajectory. It fails, however, to replicate the skew in the peak tuning of this shape, which is due to the effect of timing. In the present model motion takes as long to process as spatial position and hence the representation is only affected by motion that was present at the time of the flash, not any later. We have clearly seen that later motion does affect the

position of the flash and this would need to be incorporated at some point into the model.

We can artificially introduce a simple lag into the model. We can take the output produced above as if in fact it corresponded to a flash that had been presented 10 frames earlier and produce a graph of results (Fig. 6.22). This looks similar to those produced in Experiment 2 in Chapter 2. In this case we would match the peak of the flash representation with the motion that was actually present in the stimulus at the time of the peak, i.e. there is a 10 frame delay in the peak of the flash, but none in the calculation of the motion. A similar pattern would emerge if we leave a small e.g. 1 frame, delay for the calculation of motion. Below we illustrate the time line for a flash that appears in frame 0 at angle θ.

Input frame number	Input	Motion output	Blurred output
0	Rotating bar at angle θ	Motion centred round angle θ-10×speed	No flash, bar at position angle θ-10×speed
10	Rotating bar at angle θ+10×speed	Motion centred round angle θ	Flash peaks, bar at position angle θ

Normally, the reconstruction from frame 10 would use the motion and blurred output from frame 10 and this would correspond to a flash presented at angle θ. In Fig. 6.22 this value is said instead to correspond to a flash presented in frame θ-10. So we are using the effects of motion from the angle θ to correspond to a flash at θ-10. Or equivalently the motion at θ+10 would be taken to correspond to the flash at angle θ. So, the misalignment between two

flashes corresponds to the motion present 10 frames later than when the flashes are presented.

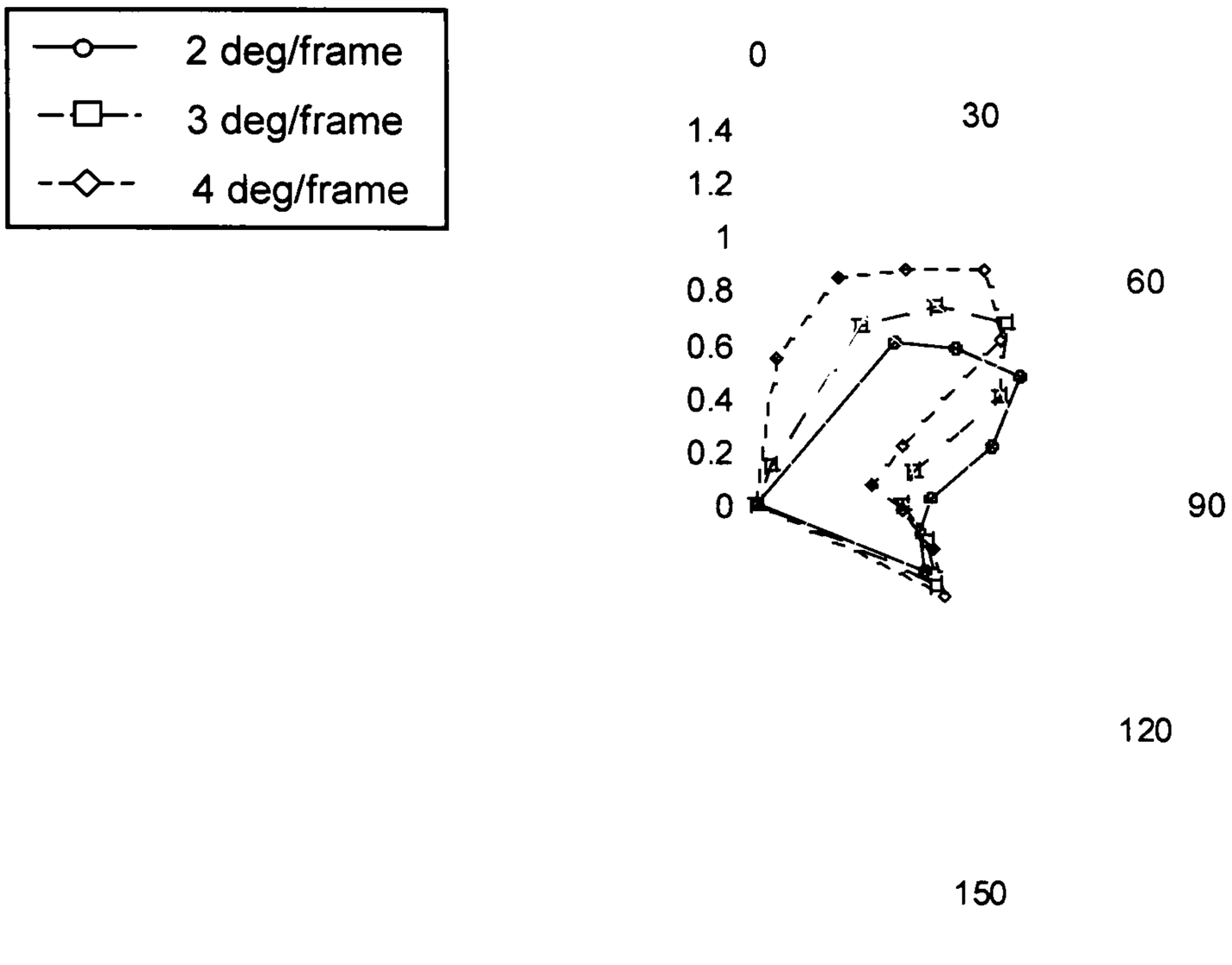


Fig. 6.22 The values from Fig. 6.21 plotted against the angle of the bar 10 frames earlier than when the flash was presented.

To actually implement this in the model we would not use the matching velocity output frame as shift parameter, but wait for some later velocity output to use as a shift parameter, skewing the calculations in time.

The problem with any such mismatch between motion and position to bear in mind is that when the spatial representation of the flash peaks, the corresponding spatial position of the bar in the blurred output will be at the same relative position as it was in the input sequence. Yet, the motion we will be using as feedback will be more advanced. If we can resolve these issues in the future, possibly by allowing the time steps in the model to be more flexible and take place within a dynamic feedback system, then we may gain an answer to this timing issue. The implication is that there is a relative delay in the flash representation as compared to the motion calculation in the biological system.

6.2.5 Conclusion from model results

In conclusion, the model presented in this chapter provides a mechanistic way of combining motion calculation with spatial representation. We have considered the effect of varying the model parameters, examining practical and biological restrictions on the implementation of the model. The model can reproduce the effect of motion influence on the position of static objects, causing oppositely drifting Gabor patches to be misaligned in the spatial representation, as well as causing small flashes presented near moving gratings to be shifted in the spatial representation. The same decrease in the influence of motion on perceived position with distance from inducing motion is shown in the model output as in experimental results. Although the original spatial reconstruction version of the model does not reflect the constant size of the shift effect w.r.t. velocity, the results can be replicated by using the model with temporal reconstruction. Finally, we see that this model output reflects to some extent the shape of the pattern of misalignment of flashes as a function of the angle of presentation in the trajectory of a moving bar, as found in Experiment 1 & 2 in Chapter 2, although only the spatial aspects of the shift are shown, with further work needed to capture the temporal dynamics of this system. Now follows a discussion of what can be learnt from this model, both from where it fails and where it succeeds. We will discuss the combined results found in the work so far and the testable predictions that can be made based on the findings.

Chapter 7- Discussion

There now follows a summary of the findings from the work presented in this thesis. First each part is summarised independently, then an overview of the combined findings and how they are linked is provided. Finally the findings are discussed in depth in relation to their implications for some of the questions raised in the introductory chapter and in light of the previous theories described.

7.1 - Summary of work

7.1.1 Empirical work

Before constructing a useful model of space-motion interactions it was necessary to further investigate some aspects of the motion induced spatial shift. In particular it was important to know whether the induced shift was specific to a region containing a moving pattern or whether it could be generated by nearby moving objects. Also it was essential for modelling work to establish whether the positional shift effect varied with distance from the inducing motion. Interesting insights were also gained by investigating how the relative timing of the presence of motion to the presentation of the flash affected the final percept.

The empirical results (Chapter 2) suggested that the relative spatial location and timing of the flashed and moving objects is important in determining the size of the apparent displacement of a briefly presented nearby static object. The finding that the effect of motion decreases with distance implies that the motion does not simply affect all briefly presented objects in the visual scene equally and that some sort of localised action is taking place. So, although long-distance interactions have been found to exist between motion and perceived position (Whitney & Cavanagh, 2000), the effect of this interaction is weaker at

further distances from motion (if not confounded with visual eccentricity). It was also found that the effect of motion could be attenuated by introducing background flicker. Typically the positional shift is observed in the direction of the nearest motion. It appears that background flicker introduces noise into this directional effect of motion, lessening the perceived shift.

The use of a motion field that varies in location, such as that generated by a moving bar, allowed an examination of how the effect of motion changed over the interval after the presentation of the flash. The size of the effect was mapped out as a function of the point along the moving bar's trajectory at which the flash was presented. Since the influence of motion increases with proximity, it was presumed that the optimal location is in the area nearest the flash (i.e. symmetrically around the horizontal). It was found that that the pattern of motion influence was not symmetric around the horizontal. The peak effect was found to be when the flash was presented around 60 ms before the moving bar reached the horizontal, and this time remained constant over different speeds of rotation of the bar. This lead to the conclusion that motion had to be present in the optimal location near an object at an optimal time *after* its presentation in order to exert maximal influence on the object's perceived location. This means that motion must influence the position of a briefly presented object over the time that the representation of the object becomes established through the neural response.

7.1.2 Modelling work

The additional results from empirical work, along with the previous results from the literature on motion induced shifts, provided the impetus for a model that would support the integration of the previously separate models of position detection and motion processing. The basis of such a model is the neural processing in V1, which in turn determines the motion calculation in V5/MT. Hence, it is necessary to look at the possible mathematical descriptions of V1 neurons' responses and the functions that can be attributed to neural systems

at this level. Evidence was provided that a good description of contrast sensitivity – a measure of V1 cell behaviour – could be provided by linearly combining differential of Gaussian functions. It was found that by adding two different orders of differentials of Gaussians with different space constants and different weights in the Fourier domain one could accurately describe unusually shaped single cell contrast sensitivity functions. This implies that, for linear cells (simple cells), receptive fields can be considered the sum of differentials of Gaussians. This model not only provided a good fit, but a simpler model with less parameters and a more intuitive explanation of the components of a V1 cell's spatial receptive field, than previous models. Sums of differentials of Gaussians can be used to build a spatial representation of the visual scene using truncated Taylor expansions (Koenderink & van Doorn, 1987) (as well as providing a useful encoding for the extraction of blur, orientation and motion) (Georgeson, 1994; Johnston et al., 1992; Sherwood & McOwan, 2003).

Frames from a sequence of images that form part of a visual scene depicting movement can be spatially represented or reconstructed using the differentials of the image produced by filtering it with derivatives of Gaussians in the x and y directions. Images are spatially sampled and luminance values are predicted in small neighbourhoods around the sampled points using the Taylor approximation. This Taylor jet representation opens up the possibility of using an already existing motion model based on this spatial representation – the Multi Channel Gradient model. Using simple stimuli we found that a sequence of images could be represented using Taylor series expansions constructed from the combined outputs of differential of Gaussian filters as calculated in the McGM.

The labile nature of this representation made it possible to introduce a spatial shift in luminance value outputs, by simply changing the weights on the outputs of filters used in the Taylor reconstruction. The values calculated by the motion model could be reintroduced into the weights of the components of the spatial representation. In this way a shift in the direction of motion was produced

between two translating Gabor pattern patches, as was found experimentally by De Valois and De Valois (1991). It was found that if motion was integrated over a larger area than each individual input into the spatial representation, then this could reproduce the effects of motion at a distance on small static, briefly presented objects. Since this model implements a local influence of motion effect, the shift decreased with distance from the inducing motion.

In order to calculate motion using the McGM (Johnston et al., 1999), the image sequence has to be blurred and differentiated in time as well as in space. By extending the reconstruction algorithm, it was also possible to reconstruct in time, for example using the derivatives in x, y and t from one image to predict the luminance values for the next image. A shift could be implemented by changing the weights on the spatial and temporal derivatives. By introducing a constant relationship between the weights of the spatial and temporal components of the Taylor reconstruction it was found that the effect of increasing velocity on the shift in spatial position was an initial increase, tending towards a constant shift with higher speeds. These methods of reconstruction allow the possibility of interpolation in space and time.

7.2 - Overall findings

Having summarised the separate results from each part of the thesis, at this point they are linked to provide an overview of the combined main findings of this work, in particular taking into consideration some of the questions raised in Chapter 1.

7.2.1 The spatial extent of the effect of motion

As mentioned above, in the experimental work in Chapter 2, it was shown that the effect of motion on perceived spatial location, although quite extensive in visual space, nevertheless is localised around the moving object and decreases with distance from the movement. It was also proposed that the action of

motion we found in the experiments in Chapter 2 was due to the larger extent of motion selective cells in V5/MT (Van Essen et al., 1981; Zeki, 1969). These ideas were further supported by the modelling work. A simple model of the effect of motion on spatial representation in which local motion estimates altered weightings in the spatial representation was generated. In this way motion can be attributed to areas of the visual scene regardless of features, making for a bottom-up approach. From this starting point, it was found that attributing motion on a strictly local point-by-point basis could explain the shift in the envelopes of moving patterns; however, it could not explain the shift of nearby briefly presented objects. As it had been suggested that the larger spatial extent of motion cells is responsible for the spatial limits of the shift (Van Essen et al., 1981; Whitney & Cavanagh, 2000) it seemed reasonable to include a spatial aggregation in the motion calculation that is not present in the spatial representation. This in fact, turned out to have the beneficial effect of improving the motion calculation, by reducing anomalies at edges without introducing implausible errors of rounding and in addition the shift in nearby objects could also be reproduced and explained.

It has been previously suggested that MT neurons may contribute to a mechanism for smoothing or averaging the velocity field (Snowden et al., 1991). The introduction of the large motion pooling areas has a similar smoothing effect on the motion field produced by the McGM. It is also interesting to note that with a motion display consisting of random moving dots, the response of MT cells saturates rapidly with dot density (Snowden et al., 1992) – something that would be predicted by our model, as all motion aggregation occurs before the ratio operation, which cancels out any multiplicative effect of higher motion density. The fact that the effect of motion appears to spread out beyond the spatial location of the object that is generating the motion in the model results, may not seem implausible in the light of the fact that a visual motion after effect can be experienced in the absence of any test visual stimulus, i.e. when viewing a blank page (Georgeson, 1976).

7.2.2 The effect of motion over time

Measuring the size of the perceived shift of briefly presented flashes either side of a rotating bar as a function of the position of the rotating bar at the time of the flash was a way of introducing a temporal aspect into the measurements of the shifts. The change in effect size as a function of relative position showed dependence on the spatial location of the moving bar, but the asymmetric peak of the effect around the horizontal suggested an added effect of timing that was not predicted by spatial proximity alone. This was further confirmed when Experiment 1 in Chapter 2 was repeated for different speeds and the skew increased, indicating that it was the *time* from the angle of peak effect to the nearest locations to the flash that was crucial. It was found that the crucial time was around 60 ms and therefore it was suggested that motion influenced the position of the flashed bar as neural response to the flashed bar became established. Although the motion feedback model does not completely address the question of temporal effects and may need to be altered to take relative neural delays into account (see Section 7.5), even the act of attempting to construct such a model focuses attention on some of the problems involved in timing in the brain. Questions we needed to ask were: if the spatial position of an object is blurred in time, taking different positions across perceptual time, which position should be taken as the perceived position? Should time be added on for the feedback process, and if so, how much? Necessarily time was built into the model due to temporal blurring. The perception of the flash is established some time later than the physical presentation of the flash, and in fact it is the motion calculated at this time that affects the flash position in the model. However some further temporal aspects may need to be considered to reproduce the data from Experiment 2 in Chapter 2.

The timing of the effect of motion could have implications for resolving the role of feedback from V5/MT. Previously the evidence for feedback has been gathered through observing the timing of figure/ground discrimination (Bullier et al., 2001; Hupé et al., 1998). However, it has been suggested that figure

ground segmentation can be implemented via lateral connections in V1 (Li, 2001). Because of the long distance effects and the effect of illusory motion, it seems less likely that motion induced shifts can be explained using lateral connections in V1. Feedback connections could be investigated using the timing of this effect.

7.2.3 The nature of spatial representation

In the data fitting and modelling work we concentrated on the questions of spatial representation raised by Experiments 1-4 in Chapter 2 and previous experiments. The fact that a flash that activates cells at a certain position on the retina can be perceived at a different location implies a dissociation at some point between retinotopic mapping and percept. V1 has been shown to have a retinotopic map (Bosking et al., 2002; Hubel & Wiesel, 1968). It has recently been shown that this retinotopy may not necessarily result in the directly corresponding pattern of activation we would expect (Whitney, 2003). Also, V1 is the primary candidate for a cortical area for accurate spatial representation as it contains a fine scale mapping (Tootell et al., 1998) (Bosking et al., 2002). This led to the investigation of how the properties of cells in this area could be used to build a representation within which these shifts could occur. We have shown that combinations of derivatives of Gaussians can provide good models for cell data and using these components for Taylor series expansions has led to a different approach to positional coding. The derivative of Gaussian description means that we can think of these cells both as spatial filters and also as signals for an accurate representation of local luminance. However, position in this framework has to be thought of as being calculated from the combined activity of cells, it is not implicit from the physical location of the cells in the cortex. Lowering the luminance at a single point does not simply decrease the firing rate of a V1 cell – for example this can depend on where in the spatial receptive field of a cell a point of light is presented (Hubel & Wiesel, 1962). Rather, it can be thought of as altering the weights on all the filters

encoding brightness, so that the combined spatial representation results in a lower value of luminance at that point.

It has been suggested in other work that the V1 retinotopic map can be thought of as flexible (Whitney, 2003). As described in Chapter 1, Section 1.4.5, the pattern of activation in the cortex was measured using fMRI, for four Gabor patterns drifting inwards towards the central fixation point or outwards away from the central fixation point (Fig. 7.6). A difference in the areas of activation in the cortex was found; however, surprisingly the inward drifting gratings had greater activation at more peripheral areas of the cortex than the outward drifting gratings. If increased activation corresponds to the perceived positions of the drifting gratings, the opposite pattern of activation would be expected. It may be more accurate to suggest that although the retinotopic mapping of connections from the retina to the striate cortex is maintained (Daniel & Whitteredge, 1961), the pattern of excitation is altered in the cortex as the representation becomes established. However this experiment also shows that the location of activation in V1, although dissociated from the stimulus also did not correlate with the final percept. This may imply a more complicated correspondence between the firing rates of individual cells and the final percept. The combined overall result may add up to a spatial representation that corresponds to the percept of a shift. When we introduce the changes in the weights we are not necessarily increasing the firing rates of the individual cells in each of the specific areas covered by the shift.

Specifically, in the feedback model at each point in space there are several filters applied to the scene corresponding to different orders of differentials of Gaussians. The insertion of the local motion value into the weights of each filter response does not necessarily increase the value produced by each filter, so we would not necessarily expect increased firing rates in the area corresponding to the new position of a shifted object.

The motion induced positional shift clearly illustrates that the problem of local sign (Lotze, 1884) is an important issue. The local jet representation was

originally suggested as an information rich way of representing the visual scene (Koenderink & van Doorn, 1987). The Taylor jet based representation is also an attempt to avoid the "neural image" idea, which fails to account for such spatial illusions. We have successfully shown how this representation can be developed to incorporate the effect of motion on position. By adopting this approach we can move away from talking about the physical distance in the input stimulus in pixels to the perceptual distance that can be deduced from the combined activity of the filter outputs in the visual cortex.

7.2.4 Mechanisms underlying motion interaction with spatial position

Much of the previous literature suggests that the spatial shift effect may be due to motion feedback from V5/MT to the primary visual cortex (Bullier, 2001a; Pascual-Leone & Walsh, 2001). The experiments in Chapter 2 strengthened this argument as the effect occurs at a distance, implying the motion calculation has to extend over large visual areas. An additional indication is that motion after the presentation of the flash carries on influencing the position of the flash. Bullier (2001b) suggested that feedback connections had an earlier effect than slower intra-cortical horizontal connections. In the motion feedback model the re-weighting of the filter outputs was implemented according to local motion calculations. The model implied that first of all the original outputs of the filters were used to calculate motion and this calculation was then re-introduced to change the output of the filters. The re-introduction of local motion estimates is the basis of a feedback connection. The use of calculated local motion estimates for the re-weighting of the components of the spatial representation provided a simple way to model the effects of motion on spatial position.

The method described does not rule out other methods of recombining the motion calculation even within the framework of re-weighting the Taylor jet based representation. For instance in the model as it stands motion is calculated frame by frame and recombined with the corresponding frame from

the spatio-temporally blurred sequence. One could argue that these frames are not necessarily matched up, for instance there could be relative delay involved between the motion calculation and the spatial representation.

Also, as we take the position of the flash in a single frame from the reconstructed output as corresponding to perceived position, this means that only one frame of motion has influence on the perceived position. Some theories of spatial localisation involve an averaging process of relative spatial position (Krekelberg & Lappe, 2000). The motion that influences the position of the flash could be averaged over frames or the position of the flash itself could be averaged over several frames.

However, if motion is extracted from the spatial filters and then re-introduced to change the spatial representation formed from the combined output of the spatial filters, then we are modelling feedback. This thesis has mostly argued in favour of a feedback explanation of the motion-induced shift, and it is not so clear on how or whether the Taylor jet representation could also be used to model the motion induced shift via lateral connections. In the present model, the motion values are used to alter the weights of the derivative filters in an additive manner. In this way the weights on the firing rates of V1 cells are linked with the firing rates of MT cells that calculate motion. These weights could alternatively be altered in proportion to the firing rates of other V1 cells. The introduction of horizontal connections in neural network models of V1 has been used to model the dynamic change in response of V1 over time to different surfaces (Li, 2001). With this strategy way V1 response might no longer correspond exactly to the physical stimulus luminance values. It is not clear how this would result in the shift of luminance values.

7.2.5 Some further implications

In the empirical work it was shown that moving objects could cause nearby static briefly presented objects to appear shifted. If spatial position is coded in the same way for moving objects as for static, then moving objects must also

be shifted ahead in the direction of their own movement. Given the way the motion shift is implemented in the model, the moving objects are indeed displaced ahead of their actual position. It is tempting to suggest that the moving objects are closer to the motion and hence shifted more than the flashed static objects, which are further from the motion. This could be proposed as the cause of the flash-lag effect, reducing it to a purely spatial extrapolation (Nijhawan, 1994). In fact, the flash-lag effect measured in Experiment 1 in Chapter 2 of 24 ms could plausibly be in the range of the shift when translated into distance at the speed of 40 rpm (about 16 min arc translation). It has been shown that surrounding motion affects the position of moving objects (Whitney & Cavanagh, 2002), but what is more difficult to ascertain is whether (and by how much) objects are shifted ahead by their own motion. However, although there have been other reports of such small flash-lag effects (Eagleman & Sejnowski, 2000), in general the effect is much larger. More importantly, the flash-lag effect increases linearly with the velocity, whereas in both cases of the motion induced shift (the translating patterns within the static envelopes and the flashes near to motion) the size of shift does not increase linearly. Hence we would have to assume the position shift increases differently with speed for a moving object than for a static object. At high speeds the flash-lag effect would result in large shifts for the moving object, far out of the range of the shifts observed for static objects (Durant & Johnston, 2004; Whitney & Cavanagh, 2000).

It seems more likely that the positional shift is a phenomenon in its own right, although it may co-exist and affect the size of the flash-lag effect. It has been suggested that the effect of motion on position may imply that the flash-lag has been under-estimated as the flash is shifted in the direction of motion (Whitney & Cavanagh, 2000). However, this ignores the possibility that the moving object itself is shifted as well and in fact, may plausibly be more shifted than the flash. If they are equally shifted, flash-lag measurements can be taken to be accurate. If the moving bar is shifted further than the flashed bar, the previous measurements have been over-estimated. Reports of the flash-lag effect

across different domains (Alais & Burr, 2003; Sheth et al., 2000) however, suggest that some sort of perceptual/cognitive difference between continuously changing and sudden changes is mostly responsible for the flash-lag effect.

When explaining any interaction of cortical areas, the time course of neural processing needs to be taken into account. Without entering into the debate of whether experience is locked into the timing of neural processes, it has been necessary to consider what takes place during the impulse response of a flash. When considering the pattern of feedback, the logical possibilities are restricted by the relative timing of activities in visual areas. It was suggested that the early peak in the size of the effect of the moving bar in Experiment 1 & 2, Chapter 2, is due to later motion affecting the position of the flash as the response to the stimulus becomes established. As has been suggested previously, feedback needs to occur more quickly than the evolution of the response in V1 in order for later motion to influence the spatial representation in V1 (Bullier, 2001b). However, this precedence does imply that motion is processed more quickly than the build-up of response in V1, and this is what causes the final pattern of results. This is not to say that the position of moving objects is processed more quickly or that we necessarily become aware of a moving object more quickly than a static flashed object. The question of how strictly neural activity in specific areas of the brain is tied in with individual percepts related to those areas is yet to be answered.

Introducing the concept of feedback as an important part of conscious awareness however (Bullier, 2001a; Pascual-Leone & Walsh, 2001), questions the idea that the relative processing time for different aspects of the scene depends on the relative evolution to suitably elevated levels of activity in spatially separate areas of the cortex. The possibility of feedback makes the relative timing of activity in areas in the brain much more complicated to disentangle and hence the timing of the emergence of a conscious percept is more difficult to try to pinpoint.

7.3 - Further questions about the model

7.3.1 Reconstruction of a scene

Do we ever reconstruct the visual scene? It has been pointed out many times that reconstructing the scene by itself is not helpful as it just leads us back to the original problem when presented with the outside world, of retrieving useful information (Dennett, 2001). However, this does not mean that the scene is not reconstructed as well as processed for relevant information. As we process visual input to extract relevant information, some detail is necessarily lost. First of all we must sample the scene according to photoreceptor spacing. The image is then filtered at the retinal ganglion level in a way that emphasises sharp changes in luminance across visual space. But are uniform areas represented at all, and if so, how? Does the brain "fill in" as it apparently does around the blind spot for example (Komatsu et al., 2002)? In the model of motion influence on spatial position developed in Chapters 4, 5 and 6, the visual scene was represented using Taylor jets, which could in theory be used to form a reconstruction, and which may perform this function of "filling in". The question remains whether the form of the representation alone is enough without an explicit reconstruction.

7.3.2 Motion capture

The model developed so far suggests that the area of estimated motion overlaps with the flashed object and it is the motion values at the same spatial position that cause the re-weighting of filter outputs to result in a shift of the briefly presented object. However, this is not a simple case of motion capture, where motion in one area of the scene affects perceived motion in a different area (Ramachandran, 1987; Ramachandran & Anstis, 1990). Whitney and Cavanagh (2000) showed that the flashes that are displaced by motion are not always perceived to be moving and hence showed that the motion induced shift is not motion capture as described.

7.3.3 Spatial interpolation

In the previous chapter it was mentioned that the space-motion interaction is necessary to implement spatial temporal interpolation, which would allow us to perceive a continuous smooth percept. In Chapter 1 the ideas of extrapolation and post-diction were introduced through the flash-lag literature. Where does this feedback mechanism lie in the context of these ideas? Perhaps it lies somewhere between the two. Where motion is smooth, to some extent the "filling in" does act as extrapolation, but in order to derive temporal gradients a sampled sequence in time after the event needs to be gathered and processed. Calculating motion and attributing it to a time in the past is indicative of a post-dictive model.

The question remains in our model of how the processes of blurring incoming information and extracting spatial position fit together in a dynamic process, rather than the step-by step picture that has been presented so far. This will be discussed in the Section 7.5.

7.3.4 Reasons for motion feedback

What is the reason for the existence of a motion-based feedback mechanism (Bullier, 2001a; Pascual-Leone & Walsh, 2001)? It is possible that it could be a verification mechanism, testing whether motion calculations are correct by checking that the spatial displacement of an object is consistent with computed motion. There has been some evidence for similar correction mechanisms that update spatial position (Arnold & Johnston, 2003). The problem with this might be that if the calculated position is "stored" in V1 then the incoming data needs to be compared against this. It is difficult to see how the two sets of data might co-exist to be compared, but not impossible.

Another possibility is that motion information is used to "fill in" temporally in the way it has been suggested that the visual system may fill in over visual space. The akinetopsic patient LM, who suffered bilateral lesions to human V5 (Zihl et

al., 1983), reported spatial position change without experienced motion. This suggests feedback might support the perception of smooth motion. Updating spatial position may be what enables us to see smooth progression between temporally sampled locations, associating motion with the relevant visual location. The lack of sensation for motion in patient LM manifested itself in a lack of smooth progression between successive spatial positions, leading to a "snapshot" version of reality. The same question emerges as before in the context of spatial representation. Do we need to reconstruct a smooth percept over time?

This idea also has some bearing on the concept of binding, i.e. how we might associate motion with a particular object in space. As we have seen, motion is calculated separately from position and more spread out over space, because of the larger size of V5/MT cell receptive fields. This feedback may be a way of returning to the accurate computation of spatial position and attributing motion to objects by updating their spatial position in between temporal samples. This would link in with the idea of spatio-temporal interpolation. It has been shown previously that temporal delay could result in spatial offset (Burr, 1979). This is an additional example of the influence of motion information on the perceived the position of an object.

7.4 - Extending the model

The sums of differentials of Gaussians function in Chapter 3 would necessarily fit all previous examples where a second order of differential of Gaussian provided an adequate fit, as this is just a sum, with one of the weighting constants set to zero. It may be useful however, to find other examples of contrast sensitivity functions of cells that were not fitted well by the second differential and see if the extended sum of differential function can provide a good fit.

It was the sums of differentials of Gaussians of different spatial scales that produced an accurate description of the contrast sensitivity functions of V1 cells in Chapter 3. For the sake of simplicity the model for spatial representation and motion calculation uses only a single scale. Work has been done on calculating motion and representing the scene at different scales, which would involve summing the differentials of different scale Gaussians. This might in future prove to be a useful extension of the present model (Dale, 2003).

The model presented in Chapter 4, 5 and 6 in certain cases it failed to predict the final percept. The most striking examples of this are a hard-edged translating pattern and a permanently presented static flash. In both these cases, although there is no reported displacement, the model predicts that there should be. This is because in the spatial aggregation of motion values the static signals disappear. In the model output there is less displacement in the direction of the surrounding motion in the case of a permanently presented flash, but nevertheless a shift does occur.

In the present model, an initial spatial representation is used to calculate motion and then the motion is fed back in to construct a new spatial representation, for a whole sequence of frames. It may be possible that the new spatial representation is used to re-calculate motion and also that the sequence of frames is re-sampled and used to update the existing spatial representation. Here are the building blocks for a dynamical system, that could settle into an equilibrium for each percept, as has been suggested previously in alternative models of V1 (Li, 2001). In this way the constant luminance values at a permanent edge or bar may contribute more strongly to the motion values over time, nulling the shift.

Although the model emulates the spatial aspect of the empirical results, it may need additional adjustments to incorporate the temporal aspect of the experiments in Chapter 2. One thing to consider is the time course of the motion calculation. The timing of processes is not inherent in the model apart from in the feedback connection. Again, this aspect may be introduced by

repeating the feedback in an iterative process that leads to an equilibrium situation, adding a temporal aspect to the calculations.

Another aspect of the calculation over time that could be further investigated is varying the extent of the time window for the reconstruction over frames. This may result in a more marked dissociation between when the motion is calculated and when the position is assigned.

Spatial information used for the motion calculation is discarded for use in the spatial reconstruction. Out of the derivatives in the 24 angular directions, only one of these directions (and the orthogonal direction) was used. A more complex, accurate description might be achieved if we combined all of these in a Taylor expansion. A full model of V1 could eventually use these to also calculate local orientation values. Some work has been done on using these filters for orientation calculation (Sherwood & McOwan, 2003).

One aspect of the model that might not be considered realistic is that the true (blurred) image luminance values are used for the positions at the centre of the spatial support area from which the Taylor expansion is calculated. There is little evidence that the visual system specifically encodes absolute luminance; rather it is relative luminance that seems important, which is then somehow anchored on a perceptual scale (Gilchrist et al.). In this case we should only use information from the first derivative and higher to represent the scene. See Fig. 7.1. Future work could look at how to arrive at a reconstruction of relative luminance using these restrictions.

Fig. 7.1 The image from Fig. 4.4. Using only derivatives from the first order and higher, the differences from each centre of expansion can be plotted, with white representing greater values than the centre of expansion and black representing less. From this we may be able to derive the final image, for instance by comparing luminance at single position to two points of expansion and building up an image of relative luminance values. Spatial blur σ=1.5, spatial blur support area = 23×23 pixels, expansion window=3×3 pixels, up to 3rd order derivatives used.

The model only reconstructs a blurred image, yet subjectively we perceive sharp images. It may be sufficient to calculate the blurriness of images as has been suggested by (Georgeson, 1994) and then attribute these values to edges. One may argue that an extra sharpening level might be necessary, in which case the scene is fully reconstructed. Scale-space models using differentials of Gaussians have been successful in reconstructing original images from sampled versions (Dale & Johnston, 2002).

Finally, it should be emphasised that the feedback model presented here is purely qualitative. We have not attempted to match the physical sizes of stimuli to the cell receptive field sizes in V1 for example. The final stages in the development in this kind of model would involve matching these properties; in this way more closely constrained parameters can be used in the model and more realistic effects can be measured. The distance moved by the bar in a given time can, for instance, be compared against the delay in the temporal

filter (Johnston & Clifford, 1995). The aggregation area of the motion calculation could also be compared against typical V5/MT receptive field size (Van Essen et al., 1981).

A more complicated structural adjustment to the model would be to introduce the scaling of receptive field sizes across eccentricities. There is existing work on mapping space onto a foveal expansion shaped map using the Taylor series (Tan et al., 2003). Hence we merely transform the rectangular grid with the same number of pixels for each filter support area to a more realistic shaped lattice, where filters at the periphery correspond to more of the image area.

It is interesting to consider the subject of re-defining the grid within which calculations are made. This might be an important notion in the case of treating the result of the model as temporal interpolation. The question is what are we interpolating between? We have seen that if we know the derivatives that correspond to an image we can reconstruct the next image in time. However this process is only possible if the temporal derivatives have already been calculated. All the original images are used for calculating the temporal derivatives in the present version of the model. It would be possible to calculate temporal derivatives using only every second image of the original sequence. Alternatively, we could reconstruct an image less than a frame ahead and in this way interpolate in time between frames.

What role would motion play in this? Using the motion to alter the weights could create a new spatial input without needing to sample the spatial position as often. In this way the spatial representation can be automatically calculated, between getting updated regularly. This would be the temporal correlate to spatial “filling in”.

7.5 - Future directions for empirical work

In the original discovery of the influence of motion on the perceived position of nearby briefly presented objects one of the findings was that the flashes appeared to have to straddle the motion for an observable misalignment to be produced. This was tested experimentally by comparing a flash presented adjacent to motion in one direction to a flash presented later at the same position, with motion in the other direction (Whitney & Cavanagh, 2000). No perceived shift was found in this case, and it was suggested that flashes have to be on either side of the motion. The model, in contrast, clearly predicts that a single flash presented on one side of a moving grating will be shifted from its original position. This experiment, however, involves a comparison of positions over time. It might be more useful to have a way of comparing position by presenting two flashes at the same time, but not straddling the motion. This way we might be able to show that central motion between two flashes is not crucial for the shift to occur.

A further interesting experiment suggested by the model and the empirical work presented here would be to present two flashes, one timed as in the original experiment and another at a slightly later time in the rotation of the bar. Would motion be observed in the presentation of the consecutive flashes? This technique, if successful, would provide a different way of measuring the differential effect of motion at different times. It would also allow an interpretation of the model's attribution of motion signals to the flash, evaluating whether the re-weighting that produces the position shift can indeed in some cases lead to a motion percept.

It has to be stated that the role of feedback has yet to be conclusively demonstrated. Localising the activation generated by illusory motion that causes a perceptual shift would show which areas are responsible for the motion that affects the shift. However, this still leaves the possibility that the position is assigned later in the chain of neural processing. That approach

would need to address the question of how fine positional information is retained and carried on into higher processing levels, whilst at the same time being aggregated for global operations on the image. In short, further experiments are needed to reveal more conclusively the role of feedback mechanisms. Trans-cranial magnetic stimulation will continue to be a useful instrument for spatially and temporally disrupting neural activity and unravelling the chain of events in the brain.

As we can see the model makes some testable predictions and raises many theoretical questions. There is also scope for extending the model to improve its performance. In future, work can be done to test the model empirically and extend the model to produce more quantitative results, so it can eventually be combined with other aspects of V1 processing apart from spatial position and motion.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America, A, 2(2), 284-299.

Alais, D., & Burr, D. (2003). The "flash-lag" effect occurs in audition and cross-modally. Current Biology, 13(1), 59-63.

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. Journal of Neurophysiology, 51(1), 16-31.

Allik, J., & Kreegipuu, K. (1998). Multiple visual latency. Psychological Science, 9, 135-138.

Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? Journal of Physiology, 97, 141-154.

Ansbacher, H. L. (1944). Distortion of perception of real movement. Journal of Experimental Psychology, 34, 1-23.

Anstis, S. (1990). Motion aftereffects from a motionless stimulus. Perception, 19(3), 301-306.

Arnold, D. H., & Clifford, C. W. G. (2002). Determinants of asynchronous processing in vision. Proceedings of the Royal Society of London: Biological Sciences, 269(1491), 579-583.

Arnold, D. H., Clifford, C. W. G., & Wenderoth, P. (2001). Asynchronous processing in vision: Colour leads motion. Current Biology, 11, 596-600.

Arnold, D. H., Durant, S., & Johnston, A. (2003). Latency differences and the flash-lag effect. Vision Research, 43, 1829-1835.

Arnold, D. H., & Johnston, A. (2003). Motion-induced spatial conflict. Nature, 425, 181-184.

Bachmann, T., & Poder, E. (2001). Change in feature space is not necessary for the flash-lag effect. Vision Research, 41, 1103-1106.

Baldo, M. V., & Klein, S. A. (1995). Extrapolation or attention shift? Nature, 378, 424.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. Journal of Physiology, 119(1), 69-88.

Barlow, H. B., & Levick, W. R. (1965). The mechanism of directional selective units in the rabbit's retina. Journal of Physiology (London), 178, 477-504.

Bevington, R. P. (1992). Data reduction and error analysis for the physical sciences (2nd ed. ed.). London: McGraw-Hill.

Borst, A., & Egelhaaf, M. (1989). Principles of visual motion detection. Trends in Cognitive Science, 12, 297-306.

Bosking, W. H., Crowley, C. C., & Fitzpatrick, D. (2002). Spatial coding of position and orientation in primary visual cortex. Nature Neuroscience, 5(9), 874-882.

Bracewell, R. N. (2000). The Fourier transform and its applications, 3rd ed. (3rd ed.). London: McGraw-Hill.

Brenner, E., & Smeets, J. B. J. (2000). Motion extrapolation is not responsible for the flash-lag effect. Vision Research, 40, 1645-1648.

Brenner, E., Smeets, J. B. J., & van den Berg, A. V. (2001). Smooth eye movements and spatial localisation. Vision Research, 41, 2253-2259.

Bruce, V., Green, P. R., & Georgeson., M. A. (1996). Chapter 8: The computation of image motion, Visual perception (physiology, psychology, and ecology) (3rd ed.). Hove: Psychology press.

Bullier, J. (2001a). Feedback connections and conscious vision. Trends in Cognitive Sciences, 5(9), 369-370.

Bullier, J. (2001b). Integrated model of visual processing. Brain Research Reviews, 36, 96-107.

Bullier, J., Hupé, J. M., James, A. C., & Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. Progress in Brain Research, 134, 193-204.

Burr, D. C. (1979). Acuity for apparent vernier offset. Vision Research, 19, 835-837.

Cai, R., & Schlag, J. (2001). Asynchronous feature binding and the flash-lag illusion. Paper presented at the The Association for Research in Vision and Opthalmology, Fort Lauderdale, Florida.

Chubb, C., & Sperling, G. (1989). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. Proceedings of the National Academy of Sciences of the United States of America, 86(8), 2985-2989.

Churan, J., & Ilg, U. J. (2002). Flicker in the visual background impairs the ability to
process a moving visual stimulus. European Journal of Neuroscience, 16, 1151-1162,.

Dale, J. L. (2003). A real-time implementation of a neuromorphic optic flow algorithm. University College London, London.

Dale, J. L., & Johnston, A. (2002). A real-time implementation of a neuromorphic optic-flow algorithm. Perception, 31 Suppl., 136c.

Daniel, P. M., & Whitteredge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. Journal of Physiology, 159, 203-221.

De Valois, R. L., Albrecht, B. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in the macaque visual cortex. Vision Research, 22, 545-599.

De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, E. D., & Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. Vision Research, 40, 3685-3702.

De Valois, R. L., & De Valois, K. K. (1990). Spatial Vision. Oxford: Oxford University Press.

De Valois, R. L., & De Valois, K. K. (1991). Vernier acuity with stationary moving Gabors. Vision Research, 31(9), 1619-1626.

De Valois, R. L., Morgan, H., & Snodderly, D. M. (1974). Psychophysical studies of monkey vision III. Spatial luminance contrat sensitivity tests of macaque and human observers. Vision Research, 14, 75-81.

Dennett, D. (2001). Are we explaining consciousness yet? Cognition, 79, 221-237.

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. Journal of Physiology, 357, 219-240.

Durant, S., & Johnston, A. (2004). Temporal dependence of local motion induced shifts in perceived position. Vision Research, 44(4), 357-366.

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. Science, 287, 2036-2028.

Eagleman, D. M., & Sejnowski, T. J. (2002). Untangling spatial from temporal illusions. Trends in Neurosciences, 25(6), 293.

Emerson, R. C., & Gerstein, G. L. (1977). Simple striate neurons in the cat. II. Mechanisms underlying directional asymmetry and directional selectivity. Journal of Neurophysiology, 40(1), 136-155.

Exner, S. (1888). Ueber optische Bewegungsempfindungen. Biologisches Centralblatt, 8, 437-448.

Ffytche, D. H., Guy, C. N., & Zeki, S. M. (1995). The parallel visual motion inputs into areas V1 and V5 of human cerebral cortex. Brain, 118, 1375-1594.

Finney, D. J. (1971). Probit analysis (3rd ed.): Cambridge University Press.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. Experimental Brain Research, 142, 139-150.

Gallant, J. L., Van Essen, D. C., & Northdurft, H. C. (1995). Two-dimensional and three-dimensional texture processing in the visual cortex of the macaque monkey. In T. Papathomas & C. Chubb & A. Gorea & E. Kowler (Eds.), Early vision and beyond (pp. 89-98). Cambridge, MA: MIT Press.

Georgeson, M. A. (1976). Antagonism between channels for pattern and movement in human vision. Nature, 259, 413-415.

Georgeson, M. A. (1980). Spatial frequency analysis in early visual processing. Philosophical Transactions of the Royal Society of London: Biological Sciences, 290(1038), 11-21.

Georgeson, M. A. (1991). Human vision combines oriented filters to compute edges. Proceedings of the Royal Society of London: Biological Sciences, 249(1326), 235-245.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. In C. F. Symposium (Ed.), Higher order processing in the visual system (Vol. 184, pp. 147-169). Chichester: Wiley.

Georgeson, M. A., & Freeman, T. C. A. (1996). Perceived location of bars and edges in one-dimensional images: Computational models and human vision. Vision Research, 37(1), 127-142.

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines: I. Quantitative studies. Journal of Experimental Psychology, 20, 453-467.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E.An anchoring theory of lightness perception. Available: http://psychology.rutgers.edu/~alan/theory3/ [2003, 8/12].

Hawken, M. J., & Parker, A. J. (1987). Spatial properties of neurons in the monkey striate cortex. Proceedings of the Royal Society of London: Biological Sciences, 231, 251-288.

Hayes, A. (2000). Apparent position governs contour-element binding by the visual sytem. Proceedings of the Royal Society of London: Biological Sciences, 267, 1341-1345.

Hess, C. V. (1904). Untersuchungen uber den Erregungsvorgang in Sehorgan bei Kurz- und bei langer daurnder Reizung. Pflugers Archiv fur die gesamte Physiologie des Menschen und der Tiere, 101, 226-262.

Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters: number, shapes and spatial covariation. Vision Research, 32(1), 47-59.

Hildreth, E., & Koch, C. (1987). The analysis of visual motion: from computational theory to neuronal mechanisms. Annual Review of Neuroscience, 10, 477-533.

Hirsch, J. A., & Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. Journal of Neuroscience, 11(6), 1800-1809.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17, 185-203.

Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. Psychonomic Bulletin and Review, 2, 322-338.

Hubel, D. H. (1995). Eye, brain and vision. New York: Scientific American Library.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular ineraction and functional architecture in the cat's visual cortex. Journal of Physiology, 160, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of the monkey striate cortex. Journal of Physiology (London), 195, 215-243.

Hubel, D. H., & Wiesel, T. N. (1974). Ordered arrangement of orientation columns in monkeys lacking visual experience. Journal of Computational Neurology, 158(3), 307-318.

Hupé, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. Journal of Physiology, 85, 134-145.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination bertween figure and background by V1, V2 and V3 neurons. Nature, 394, 784-787.

Johnston, A., & Clifford, C. W. G. (1995). A unified account of three apparent motion illusions. Vision Research, 35(8), 1109-1123.

Johnston, A., McOwan, P., & Benton, C. (1999). Robust velocity computation from a biologically motivated model of motion perception. Proceedings of the Royal Society of London: Biological Sciences, 266, 509-518.

Johnston, A., McOwan, P., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. Proceedings of the Royal Society of London: Biological Sciences, 250(1329), 297-306.

Johnston, A., & Nishida, S. (2001). Brain time or event time? Current Biology, 11, R427-R430.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 58(6), 1233-1258.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. Neuron, 15(4), 843-856.

Khurana, B., & Nijhawan, R. (1995). Extrapolation or attention shift? Nature, 378, 565-566.

Khurana, B., Watanabe, K., & Nijhawan, R. (2000). The role of attention in motion extrapolation: Are moving objects 'corrected' or flashed objects attentionally delayed? Perception, 29, 675-692.

Koenderink, J. J. (1984). Limits in Perception. AH Zeist, The Netherlands: VSP.

Koenderink, J. J., & van Doorn, A. J. (1987). Representation of local geometry in the visual system. Biological Cybernetics, 55, 367-375.

Komatsu, H., Kinoshita, M., & Murakami, I. (2002). Neural responses in the primary visual cortex of the monkey during perceptual filling-in at the blind spot. Neuroscience Research, 44, 231-236.

Krekelberg, B., & Lappe, M. (2000). A model of the perceived relative positions of moving objects based upon a slow averaging process. Vision Research, 40, 201-215.

Krekelberg, B., & Lappe, M. (2001). Neuronal latencies and the position of moving objects. Trends in Neurosciences, 24(6), 335-339.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. Journal of Neurophysiology, 16, 37-68.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregationin primary visual cortex. Journal of Neuroscience, 15(1605-1615).

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. Vision Research, 38, 2429-2454.

Li, Z. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. Neural Computation, 13, 1749-1780.

Lotze, H. (1884). Mikrokosmos. Leipzig: Hirzel.

MacKay, D. M. (1958). Perceptual stability of stroboscopically lit visual field containing self-luminous objects. Nature, 181, 507-508.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. Journal of the Optical Society of America, 70, 1297-1300.

Marr, D. (1982). Vision. New York: W. H. Freeman and Company.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. Proceedings of the Royal Society of London: Biological Sciences, 207, 187-217.

Mateeff, S., & Hohnsbein, J. (1988). Perceptual latencies are shorter for motion towards the fovea than for motion away. Vision Research, 28, 711-719.

Matin, L., Boff, K., & Pola, J. (1976). Vernier offset produced by rotary target motion. Perception and Psychophysics, 20, 138-142.

McGraw, P. V., Barrett, B. T., & Walsh, V. (2002). Motion updates perceived spatial position. Perception, 31 Suppl., 40c.

McGraw, P. V., Whitaker, D., Skillen, J., & Chung, S. T. L. (2002). Motion adaptation distorts perceived visual position. Current Biology, 12, 2042-2047.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. Journal of Neurophysiology, 55(5), 1328-1339.

Morgan, M. J., & Ward, R. M. (1985). Spatial and spatial frequency primitives in spatial-interval discrimination. Journal of the Optical Society of America, 2, 1205-1210.

Morgan, M. J., Ward, R. M., & Hole, G. J. (1990). Evidence for positional coding in hyperacuity. Journal of the Optical Society of America, A, 7(2), 297-304.

Moutoussis, K., & Zeki, S. M. (1997). A direct demonstration of perceptual asynchrony in vision. Proceedings of the Royal Society of London: Biological Sciences, 264(1380), 393-399.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in the rceptive field of simple cells in the cat's striate cortex. Journal of Physiology, 283, 53-77.

Murakami, I. (2001). A flash-lag effect in random motion. Vision Research, 41, 3101-3119.

Musseler, J., & Aschersleben, G. (1998). Localizing the first position of a moving stimulus: the Frohlich effect and an attention-shifting explanation. Perception and Psychophysics, 60(4), 683-695.

Nijhawan, R. (1994). Motion extrapolation in catching. Nature, 370(6487), 256-257.

Nijhawan, R. (1997). Visual decomposition of colour through motion extrapolation. Nature, 386, 66-69.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. Trends in Cognitive Sciences, 6(9), 387-393.

Nishida, S., & Johnston, A. (1999). Influence of motion signals on the perceived position of spatial pattern. Nature, 1999, 610-612.

Nishida, S., & Johnston, A. (2002). Marker location, not processing latency, determines temporal binding of visual attributes. Current Biology, 12(5), 359-368.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science, 292, 510-512.

Patel, S. S., Ogmen, H., Bedell, H. E., & Sampath, V. (2000). Flash-lag effect: Differential latency, not postdiction. Science, 290(5494), 1051.

Poppel, E. (1997). A hierarchical model of temporal perception. Trends in Cognitive Science, 1(2), 56-61.

Pulfrich, C. (1923). Die Stereoskopie im Dienste det isochromen und heterochromen Photometrie: Springer.

Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through visual latency. Nature, 396, 424.

Ramachandran, V. S. (1987). Interaction between colour and motion in human vision. Nature(328), 645-647.

Ramachandran, V. S., & Anstis, S. (1990). Illusory displacement of equiluminous kinetic edges. Perception, 19, 611-616.

Rao, R. P. N., Eagleman, D. M., & Sejnowski, T. J. (2000). Optimal smoothing in visual motion perception. Neural Computation, 13(6), 1243-1253.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the nervous system. In W. A. Rosenblith (Ed.), Sensory communication (pp. 303-317). New York: Wiley.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. Journal of Neurophysiology, 88, 455-463.

Rose, D. (1979). Mechanisms underlying the receptive field properties of neurons in cat visual cortex. Vision Research, 19(533-544).

Seriès, P., & Georges, S. (2002). Orientation dependent modulation of apparent speed: based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. Vision Research, 42, 2781-2797.

Sherwood, A., & McOwan, P. (2003). A biologically plausible gradient-based model for robust measures of spatial orientation. Perception(ECVP03 Supp.).

Sheth, B. R., Nijhawan, R., & Shimojo, S. (2000). Changing objects lead briefly flashed ones. Nature Neuroscience, 3(5), 489-495.

Shipp, S., de Jong, B. M., Zihl, J., Frackowiak, R. S., & Zeki, S. M. (1994). The brain activity related to residual motion vision in a patient with bilateral lesions of V5. Brain, 117(5), 1293-1302.

Shipp, S., & Zeki, S. M. (1989). The organization of connections between areas V5 and V1 in macaque monkey visual cortex. European Journal of Neuroscience, 1, 308-331.

Snowden, R. J. (1998). Shifts in perceived position following adaptation to visual motion. Current Biology, 8, 1343-1345.

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. Experimental Brain Research, 88(2), 389-400.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. Journal of Neuroscience, 11(9), 2768-2785.

Stark, J. (1997). Adaptive model selection using orthogonal least squares methods. Proceedings of the Royal Society of London, 453, 21-42.

Sullivan, G. D., Oatley, K., & Sutherland, N. S. (1972). Vernier acuity as affected by target length and separation. Perception and Psychophysics, 12, 138-444.

Tan, S., Dale, J. L., & Johnston, A. (2003). Performance of three recursive algorithms for fast space-variant Gaussian filtering. Real Time Imaging, 9, 215-228.

Tootell, R. B., Hadjikhani, W., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. Proceedings of the National Academy of Sciences of the United States of America, 95(3), 811-817.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., & Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. Journal of Neuroscience, 15(4), 3215-3130.

van de Grind, W. (2002). Physical, neural, and mental timing. Consciousness and Cognition, 11(2), 241-264; discussion 308-213.

Van Essen, D. C., Maunsell, J. H., & Bixby, J. L. (1981). The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. Journal of Comparative Neurology, 199(3), 293-326.

Watanabe, K., Nijhawan, R., & Shimojo, S. (2002). Shifts in perceived position of flashed stimuli by illusory object motion. Vision Research, 42, 2645-2650.

Watson, A. B. (1983). Detection and recognition of simple spatial forms. In O. J. Braddick & A. C. Sleigh (Eds.), Physical and biological processing of images (pp. 100-114). Berlin: Springer-Verlag.

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., & Zeki, S. M. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cerebral Cortex, 3(2), 79-94.

Wenderoth, P., & Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. Vision Research, 28(2), 301-312.

Westheimer, G. (1981). Visual hyperacuity, Progress in sensory physiology: Vol. 1. Berlin: Springer-Verlag.

Whitney, D. (2002). The influence of visual motion on spatial position. Trends in Cognitive Sciences, 6(5), 211-216.

Whitney, D. (2003). Flexible retinotopy: motion-dependent position coding in the visual cortex. Science, 302(5646), 878-881.

Whitney, D., & Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived position of remote stationary objects. Nature Neuroscience, 3(9), 954-959.

Whitney, D., & Cavanagh, P. (2002). Surrounding motion affects the perceived locations of moving stimuli. Visual Cognition, 9, 139-152.

Whitney, D., Cavanagh, P., & Murakami, I. (2000). Temporal facilitation of moving stimuli is independent of changes in direction. Vision Research, 40, 3829-3839.

Whitney, D., & Murakami, I. (1998). Latency difference, not motion extrapolation. Nature Neuroscience, 1(8), 656-657.

Wilson, H., & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America, A, 1, 124-131.

Yang, J., Xiaofeng, Q., & Makous, M. (1995). Zero frequency masking and a model of contrast sensitivity. Vision Research, 35(14), 1965-1978.

Young, R. A. (1985). The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles. General Motors Research, 4920.

Young, R. A. (1986). The Gaussian derivative model for machine vision: Visual cortex simulation. General Motors Research, 5323(1-24).

Zanker, J. M. (1994). Modeling human motion perception 1. Classical stimuli. Naturwissenschaften, 81, 156-163.

Zanker, J. M., Quenzer, T., & Fahle, M. (2001). Perceptual deformation induced by visual motion. Naturwissenschaften, 88(3), 129-132.

Zanker, J. M., Srinivasan, M. V., & Egelhaaf, M. (1999). Speed tuning in elementary motion detectors of the correlation type. Biological Cybernetics, 80(2), 109-116.

Zeki, S. M. (1969). Representation of central fields in prestriate cortex of monkey. Brain Research, 19, 63-75.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. Nature, 274(5670), 423-428.

Zeki, S. M., & Bartels, A. (1999). Toward a theory of visual consciousness. Consciousness and Cognition, 8(2), 225-259.

Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. Brain, 106 (pt 2), 313-340.